

MOMENTS AND CUMULANTS FROM GENERATING
FUNCTIONS OF HILBERT SPACE-VALUED
RANDOM VARIABLES AND AN APPLICATION
TO THE WISHART DISTRIBUTION

by

Björn Holmquist

1985:3



UNIVERSITY OF LUND AND
LUND INSTITUTE OF TECHNOLOGY
DEPARTMENT OF MATHEMATICAL STATISTICS

LUND, SWEDEN

TYP AV DOKUMENT ☒ Doktorsavhandling ☐ Examensarbete ☐ Kompendium		☐ Ansökan ☐ Reserapport ☐ Delrapport ☐ Slutrapport	☐ Tidskriftsartikel ☐ Konferensuppsats ☐	DOKUMENTBETECKNING/CODEN LUTFD2/(TFMS-3037)/1-86/(1985)
AVDELNING/INSTITUTION Department of Mathematical Statistics Box 118 S-221 00 Lund, Sweden				
FÖRFATTARE Björn Holmquist				
DOKUMENTTITEL OCH UNDERTITEL MOMENTS AND CUMULANTS FROM GENERATING FUNCTIONS OF HILBERT SPACE-VALUED RANDOM VARIABLES AND AN APPLICATION TO THE WISHART DISTRIBUTION				
SAMMANFATTNING The procedure for deriving moments and cumulants by means of differentiating generating functions, is investigated for random variables in Hilbert spaces. Relations are given between moments and cumulants considered as Hilbert space elements. We consider different choices of coordinate space representations of differentials and of moments and cumulants which are related in the sense that the representations of moments and cumulants are obtained from representations of differentials of generating functions. As an example of the technique we derive, in matrix formulation, the moments and cumulants of the noncentral Wishart distribution over real symmetric matrices.				
NYCKELORD Moment, cumulant, polynomial transformation, noncentral Wishart distribution, vec operator, direct product permuting matrices, direct product transposition matrix, commutation matrix, permutation operator.				
DOKUMENTTITEL OCH UNDERTITEL – SVENSK ÖVERSÄTTNING AV UTLÄNDSK ORIGINALTITEL				
TILLÄMPNINGSOMRÅDE				
TILLÄMPNINGSNYCKELORD				
UTGIVNINGSDATUM år 1985 mån 04		ANTAL SID 86		SPRÅK ☐ svenska ☒ engelska ☐ annat
ÖVRIGA BIBLIOGRAFISKA UPPGIFTER				ISSN
				ISBN
				PRIS
I, the undersigned, being the copyright owner of the abstract, hereby grant to all reference source permission to publish and disseminate the abstract.				
Date 1985-04-29				Signature Björn Holmquist

CONTENTS

	Page
Part I DERIVATION OF MOMENTS AND CUMULANTS FROM GENERATING FUNCTIONS OF RANDOM VARIABLES IN HILBERT SPACES	
1. INTRODUCTION	1
2. NOTATIONS AND PRELIMINARIES	4
3. SOME DIFFERENTIALS ON BANACH SPACES	12
4. MOMENTS AND CUMULANTS FROM DIFFERENTIALS OF GENERATING FUNCTIONS	21
4.1. Introduction	21
4.2. Moments and cumulants of Hilbert space-valued random variables	21
4.3. Relations between moments and cumulants	25
4.4. Some properties	35
Part II DERIVATION OF MOMENTS AND CUMULANTS FROM GENERATING FUNCTIONS OF RANDOM VARIABLES IN COORDINATE SPACES	
5. COORDINATE SPACE DIFFERENTIALS: GRADIENTS	41
5.1. Introduction	41
5.2. General coordinate space differentials	41
5.3. Matrix derivatives	44
5.4. Symmetric matrix derivatives	57
5.5. Matrix derivatives with linear structure	65
6. COORDINATE SPACE MOMENTS AND CUMULANTS: MOMENT AND CUMULANT ARRAYS	69
Part III AN APPLICATION IN THE SPACE OF SYMMETRIC REALVALUED MATRICES	
7. THE CUMULANTS AND MOMENTS OF THE NON-CENTRAL WISHART DISTRIBUTION OVER REAL SYMMETRIC MATRICES	72
7.1. Introduction	72
7.2. Notations	72
7.3. Some matrix derivatives	75
7.4. Cumulants and moments of the Wishart distribution	80
REFERENCES	84

1. INTRODUCTION

The procedure for deriving moments of scalar random variables from the moment generating function or characteristic function is well-known. Since the paper of Thiele (1899), a similar procedure for deriving cumulants of scalar random variables by means of differentiating the cumulant generating function or the second characteristic is also well-known. When generalizing these procedures to vector random variables, special considerations have to be taken to the following circumstances.

(i) In the scalar case there is generally no need to distinguish between an element of the field and a linear or multilinear mapping from the field to the field. The derivative at zero of a generating function (which is a linear or multilinear mapping) can thus also be considered as an element of the specific field. While in the scalar case moments and cumulants can thus be obtained as derivatives at zero of the generating functions, considerations have to be made in the vector case of the domain and range spaces of the linear or multilinear mappings obtained by differentiating the generating functions. Distinction has to be made between the differentials of the generating functions and the moments and cumulants considered as elements of some vector space and the connection between them has to be obtained.

(ii) For random vector variables which have representations in a coordinate space, we wish to have a coordinate space representation of the moments as well as of the cumulants, and these representations are then to be considered as elements of some coordinate space.

(iii) For coordinate space representations of random variables the generating functions have coordinate space representations of the arguments. A method for obtaining coordinate representations of moments and cumulants from these representations of generating functions has to be developed by developing coordinate space representations of derivatives.

Such a technique will then allow us to obtain coordinate representations of moments and cumulants by a direct differentiation of the representations of generating functions, not coordinate-wise but as an element in a coordinate space.

In part I of the paper we work out the procedure for random variables in Hilbert spaces. The development is held in a coordinate free vector space approach but we also show how coordinate representations are obtained.

Section 2 is concerned with basic results on tensor spaces and completely symmetric spaces over a Hilbert space. We give some preliminary results for a class of symmetrizers on tensor spaces which play an important role in the development.

The task of section 3 is to give basic results for differentials of mappings with domain and range in Banach spaces. Differentials needed for the development in section 4 are here also given for certain mappings on Hilbert spaces.

In section 4 we define moments and cumulants of random variables in Hilbert spaces by means of mappings of differentials of generating functions. We show how this definition induces a natural extension to vector spaces of the definition of moments of scalar random variables as expectations of monomials of the random variable. We also generalize the relations between moments and cumulants of scalar random variables to the vector space case.

In part II of the paper we discuss the results for separable Hilbert spaces. Especially we discuss the aspect of different coordinate space representations of moments and cumulants and their relations to coordinate space representations of differentials. The technique for obtaining realizations of moments and cumulants by using realizations of differentials of generating functions is emphasized, not the existence differentials, moments or cumulants.

In section 5 we give examples of realizations of differentials of mappings between coordinate spaces. Special emphasis is laid on the action of the representations on their arguments.

In section 6 we give examples of representations of moments and cumulants of vector valued random variables by means of isomorphic mappings into coordinate spaces.

In part III, consisting of section 7, we apply - as an example of the technique - the procedure to the moment generating function of a Wishart distributed random matrix. We derive the moments and cumulants of arbitrary order of this matrix.

2. NOTATIONS AND PRELIMINARIES.

Denote by $M(X;V)$ and $L(X;V)$ the space of general mappings and the space of linear mappings from X to V , respectively. If $f \in M(X;V)$ then $f(x)$ or fx denotes the image of $x \in X$ under f and for $V \subset X$ we let $f|V$ denote the restriction to V of f , that is, $f|V \in M(V;V)$ is such that $(f|V)(x) = f(x)$ for $x \in V$. For $D \subset X$ we let $f(D)$ or fD denote the set $\{f(x), x \in D\}$. When $g \in M(V;W)$ we denote the composition of f and g by $g \circ f \in M(X;W)$. The constant mapping with value y is denoted by c_y , that is, if $c_y \in M(X;V)$ then $c_y(x) = y \in V$ for all $x \in X$.

Let X and V be inner product spaces over the same field K (the field R of real numbers or the field C of complex numbers) with inner products $\langle \cdot, \cdot \rangle_X$ and $\langle \cdot, \cdot \rangle_V$, respectively. The inner products will be considered as linear in the first variable, that is, for an arbitrary inner product $\langle \cdot, \cdot \rangle$ and $a_1, a_2, b_1, b_2 \in K$,

$$\begin{aligned} \langle a_1 x_1 + a_2 x_2, b_1 y_1 + b_2 y_2 \rangle &= a_1 \alpha(b_1) \langle x_1, y_1 \rangle + a_1 \alpha(b_2) \langle x_1, y_2 \rangle \\ &\quad + a_2 \alpha(b_1) \langle x_2, y_1 \rangle + a_2 \alpha(b_2) \langle x_2, y_2 \rangle, \end{aligned}$$

where α is the involutory automorphism on K in the definition of $\langle \cdot, \cdot \rangle$.

The identity map on X is denoted $I_X \in L(X;X)$. We use the notations 1_K and 0_K for the unit element and the zero element in K , respectively and denote by δ_{ij} Kronecker's delta which is 1_K if $i=j$ and 0_K if $i \neq j$.

The adjoint of $A \in L(X;V)$ is denoted by $A^* \in L(V;X)$, that is,

$$\langle A^* y, x \rangle_X = \langle y, Ax \rangle_V \quad x \in X, y \in V.$$

and we write X^* for the dual space $L(X;K)$ of X . For a reflexive inner product space X , we write an arbitrary mapping $f \in X^*$ as x_0^* when it is considered as the adjoint of $x_0 \in L(K;X)$ defined by $x_0(k) = kx_0$, $k \in K$, where x_0 on the right hand side is an element in X . This implies that

$$\langle x, x_0(k) \rangle_X = \langle x_0^*(x), k \rangle_K = x_0^*(x) \alpha(k) \langle 1_K, 1_K \rangle_K$$

and thus

$$(2.1) \quad x_0^*(x) = \langle x, x_0 \rangle_X / \langle 1_K, 1_K \rangle_K$$

for all $x \in X$. The element $x_0 \in X$, such that $f(x) = x_0^*(x)$ for a specified $f \in X^*$ and for all $x \in X$, is unique and is given by

$$(2.2) \quad x_0 = \sum_{i \in J} f(x_i^b) x_i^b / \langle 1_K, 1_K \rangle_K$$

for any orthogonal basis $\{x_i^b, i \in J\}$ of X . We do not distinguish between the element $x_0 \in X$ and the mapping $x_0 \in L(K, X)$ and call x_0^* the adjoint of $x_0 \in X$.

For $A \in L(X; Y)$ we define $\|A\| = \sup\{\|Ax\|_Y; \|x\|_X \leq 1\}$ where $\|\cdot\|_X$ is the norm given by the inner product $\langle \cdot, \cdot \rangle_X$ on X .

Let $X_i, i=1, \dots, r$ be finite dimensional inner product spaces over the same field K with inner products $\langle \cdot, \cdot \rangle_{X_i}, i=1, \dots, r$. The space of r -linear mappings from $X_1 \times \dots \times X_r$ to Y is denoted by $L(X_1, \dots, X_r; Y)$ or, when $X_1 = \dots = X_r = X$, by $L({}^rX; Y)$. We denote by $\otimes^r$ a specific r -linear mapping with domain on $X_1 \times \dots \times X_r$ such that, with $X_1 \otimes \dots \otimes X_r$ denoting the linear closure of the range of $\otimes^r$, the pair $(X_1 \otimes \dots \otimes X_r, \otimes^r)$ is a tensor product of $X_1, \dots, X_r$. We write $X^{<r>} (X^{<0>} = K, X^{<1>} = X)$ for $X_1 \otimes \dots \otimes X_r$ when $X_1 = \dots = X_r = X$. A decomposable tensor in $X_1 \otimes \dots \otimes X_r$ is denoted by

$$x_1 \otimes \dots \otimes x_r \quad \text{or} \quad \bigotimes_{i=1}^r x_i$$

when written as a tensor product of the vectors $x_i \in X_i, i=1, \dots, r$. We use the notation $x^{<r>} (x^{<0>} = 1_K, x^{<1>} = x)$ for $\bigotimes_{i=1}^r x$, the r 'th ($r=0, 1, \dots$) tensor power of $x \in X$ and denote by $\pi_X^r \in M(X, X^{<r>})$ the tensor power mapping, $\pi_X^r(x) = x^{<r>}, x \in X$, on X .

The inner product under consideration in $X_1 \otimes \dots \otimes X_r$ will be the one naturally induced by the inner products in $X_i, i=1, \dots, r$, and

which satisfies

$$(2.3) \quad \langle x_1 \otimes \dots \otimes x_r, y_1 \otimes \dots \otimes y_r \rangle_{X_1 \otimes \dots \otimes X_r} = \prod_{i=1}^r \langle x_i, y_i \rangle_{X_i}$$

for $x_i, y_i \in X_i, i=1, \dots, r$.

If $A_i \in L(X_i; Y_i), i=1, \dots, r$, then $A_1 \otimes \dots \otimes A_r$ or $\bigotimes_{i=1}^r A_i$ denotes the transformation naturally induced in $L(X_1 \otimes \dots \otimes X_r; Y_1 \otimes \dots \otimes Y_r)$, by

$$\left(\bigotimes_{i=1}^r A_i \right) \left(\bigotimes_{i=1}^r x_i \right) = \bigotimes_{i=1}^r A_i x_i$$

for $x_i \in X_i, i=1, \dots, r$.

Given a complex C of elements of a group G , a character χ of degree 1 on G and a homomorphism $A \in M(G; GL(V))$ of G , $GL(V)$ being the general linear group on V , we define a symmetrizer $S_C^{\chi, A}$ on V as the linear transformation given by

$$S_C^{\chi, A} = |C|^{-1} \sum_{c \in C} \chi(c) A(c),$$

where $|C|$ is the cardinal of C . For two characters θ and χ of degree 1 on G we define $\langle \theta, \chi \rangle_C$ by

$$\langle \theta, \chi \rangle_C = \sum_{c \in C} \theta(c) \chi(c^{-1}).$$

For a linear space V and $A_i \in L(V, V), i=1, \dots, r$, we denote by $V_1(A_1, \dots, A_r)$ the linear subspace of V of eigenvectors to $A_1, \dots, A_r$ defined by

$$V_1(A_1, \dots, A_r) = \{x \in V; A_i x = x, i=1, \dots, r\}.$$

A projection operator on $V \subset X$ is denoted P_V .

The results given in the following theorem are known, cf. e.g. Marcus (1973, p. 78), and are stated here for easy reference.

THEOREM 2.1.

$$(i) \quad S_C^{\theta, A} \circ S_G^{\chi, A} = \langle \theta, \chi \rangle_C |C|^{-1} S_G^{\chi, A} = S_G^{\chi, A} \circ S_C^{\theta, A}.$$

(ii) If G is Abelian or either of $\bar{\theta}$ and χ is c_1 or $\theta=\chi$ then

$$\langle \theta, \chi \rangle_G = \delta_{\theta, \chi} |G|.$$

$$(iii) \quad S_G^{\chi, A} = P_{Y_1}(\{\chi(g)A(g); g \in G\}).$$

(iv) If $A(g)$ is unitary for every $g \in G$, then

$$S_G^{\chi, A} = (S_G^{\chi, A})^*.$$

PROOF. (i): If $c \in C$ then

$$\begin{aligned} \theta(c)A(c) \circ S_G^{\chi, A} &= \theta(c) |G|^{-1} \sum_{g \in G} \chi(g)A(c)A(g) \\ &= \theta(c)\chi(c^{-1}) |G|^{-1} \sum_{g \in G} \chi(cg)A(cg) = \theta(c)\chi(c^{-1}) S_G^{\chi, A}. \end{aligned}$$

By summing over $c \in C$ and dividing by $|C|$ we obtain the first equation of (i). A similar argument yields the second equation.

(ii): Multiply $\langle \theta, \chi \rangle_G$ by $\theta(g)\chi(g^{-1})$, $g \in G$ and sum over G . Then

$$\begin{aligned} (\langle \theta, \chi \rangle_G)^2 &= \sum_{g \in G} \sum_{h \in G} \theta(g)\chi(g^{-1})\theta(h)\chi(h^{-1}) \\ &= \sum_{g \in G} \sum_{h \in G} \theta(gh)\chi((hg)^{-1}) = |G| \sum_{f \in G} \theta(f)\chi(f^{-1}) \end{aligned}$$

where the last equation follows by the commutative property $gh = hg = f$ or if either of θ and χ is c_1 . Hence $(\langle \theta, \chi \rangle_G)^2 = |G| \langle \theta, \chi \rangle_G$ and so $\langle \theta, \chi \rangle_G$ is $|G|$ or 0 . If $\langle \theta, \chi \rangle_G = |G|$, then since $\theta(g)\chi(g^{-1}) \langle \theta, \chi \rangle_G = \langle \theta, \chi \rangle_G$, $\theta(g)\chi(g^{-1}) = 1$ and thus $\theta(g) = \chi(g)$ for all $g \in G$.

(iii): By taking $C = G$ and $\theta = \chi$ in (i) we see that $S_G^{\chi, A}$ is idempotent. Further, by taking $C = \{g\}$ and $\theta = \chi$ in (i) we have

$$(\chi(g)A(g) \circ S_G^{\chi, A})(y) = S_G^{\chi, A}(y), \quad g \in G, y \in Y$$

and hence $S_G^{\chi, A}(y) \in Y_1(\{\chi(g)A(g); g \in G\})$.

(iv): We have

$$\begin{aligned} \langle y, S_G^{\chi, A} x \rangle &= \langle y, |G|^{-1} \sum_{g \in G} \chi(g)A(g)x \rangle = \alpha(|G|^{-1}) \sum_{g \in G} \alpha(\chi(g)) \langle y, A(g)x \rangle \\ &= \alpha(|G|^{-1}) \sum_{g \in G} \alpha(\chi(g)) \langle A(g)^* y, x \rangle \\ &= \langle \alpha(|G|^{-1}) \sum_{g \in G} \alpha(\chi(g)) A(g)^* y, x \rangle \end{aligned}$$

$$= \langle |G|^{-1} \sum_{g \in G} \chi(g^{-1}) A(g^{-1}) y, x \rangle = \langle S_G^{\chi, A} y, x \rangle.$$

where the last step follows from that $\alpha(\chi(g)) = \chi(g^{-1})$ for a character of degree 1, $\alpha(|G|^{-1}) = |G|^{-1}$ and $A(g)^* = A(g)^{-1} = A(g^{-1})$. $\square$

Let $\underline{r}_k = (r_1, \dots, r_k)$ be a partition of weight k of the non-negative integer $r = \sum_{i=1}^k r_i$ into non-negative integer parts. We write $\begin{pmatrix} r \\ \underline{r}_k \end{pmatrix}$ for $r! / \prod_{i=1}^k r_i!$ and $\begin{pmatrix} p \\ \underline{r}_k \end{pmatrix}$ for $p! / ((p-r)! \prod_{i=1}^k r_i!)$. The partition is positive if $r_1, \dots, r_k$ are all positive. Denote by $\underline{1}_r$ the complete positive partition of r , that is $\underline{1}_r = (1, \dots, 1)$ (r times 1), and by $S_{\underline{1}_r}$ the symmetric group of degree r . Let $A_r \in M(S_{\underline{1}_r}; GL(X_r))$ with $\dim(X_r) = r$, for $r=1, 2, \dots$ be a sequence of homomorphisms of $S_{\underline{1}_r}$. Let in addition these be such that for every partition $\underline{r}_k$ of r and every $\xi_i \in S_{\underline{1}_{r_i}}$, $i=1, \dots, k$, there exist $\sigma \in S_{\underline{1}_r}$ such that

$$A_{r_1}(\xi_1) \otimes \dots \otimes A_{r_k}(\xi_k) = A_r(\sigma)$$

where $A_0(\xi_j) = 1_K$ if $\xi_j \in S_{\underline{1}_0}$. Clearly σ is uniquely determined by $\xi_1, \dots, \xi_k$ and we write $\sigma = \xi_1 \otimes \dots \otimes \xi_k$.

Define the relation $\stackrel{\underline{r}_k}{=}$ by letting $\xi \stackrel{\underline{r}_k}{=} \sigma$, for $\xi, \sigma \in S_{\underline{1}_r}$, if there exist $\eta_i \in S_{\underline{1}_{r_i}}$, $i=1, \dots, k$, such that $\sigma^{-1}\xi = \eta_1 \otimes \dots \otimes \eta_k$. This equivalence relation (which is easily verified) partitions $S_{\underline{1}_r}$ into $\begin{pmatrix} r \\ \underline{r}_k \end{pmatrix}$ equivalence classes, each of which contains $\prod_{i=1}^k r_i!$ elements. We denote by $S_{\underline{r}_k}$ any complex which consists of one element from each of these equivalence classes. Clearly $\stackrel{\underline{1}_r}{=}$ is identical with the ordinary equality.

Let $\sigma \in S_{\underline{1}_r}$ be given by

$$(2.4) \quad \sigma = \begin{pmatrix} 1 & 2 & \dots & r \\ \sigma^{-1}(1) & \sigma^{-1}(2) & \dots & \sigma^{-1}(r) \end{pmatrix}$$

where $\sigma^{-1}(1), \dots, \sigma^{-1}(r)$ are the symbols $1, \dots, r$ in some order. We denote by $P_r^{\otimes}(\sigma)$ the permutation operator on $X^{<r>}$ associated with σ ,

that is, the unique linear mapping in $GL(X^{<r>})$ satisfying

$$P_r^{\otimes(\sigma)}(x_1 \otimes \dots \otimes x_r) = x_{\sigma^{-1}(1)} \otimes \dots \otimes x_{\sigma^{-1}(r)}$$

for all $x_i \in X$, $i=1, \dots, r$.

When $y = X^{<r>}$, $G = S_{\underline{1}_r}$, $C = S_{\underline{r}_k}$, $\chi = c_1$ and $A = P_r^{\otimes}$ for a partition $\underline{r}_k$ of r , we write $S_{\underline{r}_k}$ for the symmetrizer $S_C^{\chi, A}$, that is,

$$(2.5) \quad S_{\underline{r}_k} = \sum_{\sigma \in S_{\underline{r}_k}} P_r^{\otimes(\sigma)} \left(\prod_{i=1}^k r_i! \right) / r!$$

It follows from the definition of $S_{\underline{r}_k}$ that

$$\left(\begin{matrix} r \\ \underline{r}_k \end{matrix} \right) S_{\underline{r}_k} (x_1^{<r_1>} \otimes \dots \otimes x_k^{<r_k>})$$

is the sum over all distinct terms of type $y_1 \otimes \dots \otimes y_r$ for which r_1 of the y_i 's are equal to x_1 , r_2 of the y_i 's are equal to x_2 etc. The sum contains $\left(\begin{matrix} r \\ \underline{r}_k \end{matrix} \right)$ such terms. By summing over all partitions of r of weight k we obtain the following multinomial expansion of tensor powers

$$(2.6) \quad \left(\sum_{i=1}^k x_i \right)^{<r>} = \sum_{\underline{r}_k} \left(\begin{matrix} r \\ \underline{r}_k \end{matrix} \right) S_{\underline{r}_k} (x_1^{<r_1>} \otimes \dots \otimes x_k^{<r_k>}).$$

We shall need the following result concerning partitions $(\underline{1}_s, r-s)$ of r for various s , $1 \leq s \leq r-1$. We use the notation $lc\{x_i \in X, i \in I\}$ for the smallest linear subspace of X containing $\{x_i \in X, i \in I\}$.

LEMMA 2.1. The mappings $S_{\underline{1}_s, r-s} \circ (S_1^{<s>} \otimes S_{1, r-s-1})$ and $S_{\underline{1}_{s+1}, r-s-1}$ coincide on $lc\left\{ \left(\bigotimes_{i=1}^{s+1} x_i \right) \otimes x_{s+2}^{<r-s-1>}, x_i \in X, i=1, \dots, s+2, \right\}$ for $s=0, \dots, r-1$.

PROOF. Let $\hat{S}_{\underline{r}_k} = \left(\begin{matrix} r \\ \underline{r}_k \end{matrix} \right) S_{\underline{r}_k}$. Then

$$\begin{aligned} (\hat{S}_1^{<s>} \otimes \hat{S}_{1, r-s-1}) \left(\left(\bigotimes_{i=1}^{s+1} x_i \right) \otimes x_{s+2}^{<r-s-1>} \right) \\ = \left(\bigotimes_{i=1}^s x_i \right) \otimes (\hat{S}_{1, r-s-1} (x_{s+1} \otimes x_{s+2}^{<r-s-1>})) \end{aligned}$$

Here $\hat{S}_{1, r-s-1} (x_{s+1} \otimes x_{s+2}^{<r-s-1>})$ is a sum of the $r-s$ terms

$$x_{s+2}^{<j>} \otimes x_{s+1} \otimes x_{s+2}^{<r-s-1-j>}, j=0,1,\dots,r-s-1.$$

We note that $\hat{S}_{\underline{1}_s, r-s}((\bigotimes_{i=1}^s x_i) \otimes x^{<r-s>})$ is the sum of the $\begin{pmatrix} r \\ \underline{1}_s, r-s \end{pmatrix}$ distinct terms of type $y_1 \otimes \dots \otimes y_r$ where $r-s$ of the y_i 's are equal to x and the remaining ones are equal to $x_1, \dots, x_s$. Hence it follows that

$$\hat{S}_{\underline{1}_s, r-s}((\bigotimes_{i=1}^s x_i) \otimes (\hat{S}_{\underline{1}, r-s-1}(x_{s+1} \otimes x_{s+2}^{<r-s-1>})))$$

is equal to the sum of the $\begin{pmatrix} r \\ \underline{1}_s, r-s \end{pmatrix} (r-s) = \begin{pmatrix} r \\ \underline{1}_{s+1}, r-s-1 \end{pmatrix}$ distinct terms of type $y_1 \otimes \dots \otimes y_r$ where $r-s-1$ of the y_i 's are equal to x_{s+2} and the rest are equal to $x_1, \dots, x_{s+1}$. But this is also equal to

$$\hat{S}_{\underline{1}_{s+1}, r-s-1}((\bigotimes_{i=1}^{s+1} x_i) \otimes x_{s+2}^{<r-s-1>}).$$

This shows that $\hat{S}_{\underline{1}_s, r-s} \circ (\hat{S}_1^{<s>} \otimes \hat{S}_{\underline{1}, r-s-1})$ and $\hat{S}_{\underline{1}_{s+1}, r-s-1}$ coincide on the space generated by decomposable tensors of type

$$(\bigotimes_{i=1}^{s+1} x_i) \otimes x_{s+2}^{<r-s-1>}, x_i \in X, i=1, \dots, s+2,$$

and hence also on the linear closure of this space. By dividing with

$$\begin{pmatrix} r \\ \underline{1}_{s+1}, r-s-1 \end{pmatrix} = \begin{pmatrix} r \\ \underline{1}_s, r-s \end{pmatrix} \begin{pmatrix} r-s \\ \underline{1}, r-s-1 \end{pmatrix} \quad \text{the result follows.} \quad \square$$

We denote by $L({}^rX; V; G; \chi)$ the space of all r -linear mappings which are symmetric with respect to a subgroup G of $S_{\underline{1}_r}$ and the character χ of degree 1 on G , that is, if $\phi \in L({}^rX; V; G; \chi)$ then

$$\phi(x_{\sigma^{-1}(1)}^{<1>}, \dots, x_{\sigma^{-1}(r)}^{<r>}) = \chi(\sigma) \phi(x_1, \dots, x_r)$$

for all $\sigma \in G$ and all $x_i \in X, i=1, \dots, r$. We write $\bigodot({}^rX; V)$ for $L({}^rX; V; S_{\underline{1}_r}; c_1)$; the space of completely symmetric r -linear mappings.

We denote by $X^{(r)}$ the r 'th completely symmetric space over X , that is, the symmetry class of tensors associated with $S_{\underline{1}_r}$ and $\chi \equiv 1$. By $\bigodot({}^rX; X^{(r)})$ we mean the specific completely symmetric r -linear mapping in $L({}^rX; X^{(r)})$ which is related to $\bigotimes^r$ by

$$\odot^r = s_{1-r} \circ \otimes^r.$$

A decomposable element in $X^{(r)}$ (which is the linear closure of the range of $\odot^r$) is denoted $x_1 \odot \dots \odot x_r$ or $\bigodot_{i=1}^r x_i$ when written as a symmetric product of the vectors $x_i \in X, i=1, \dots, r$.

We mention the well-known fact that if $A \in L(X; Y)$ and $B \in L(Y; W)$, then $B \circ A$ can be regarded as the image of B under the linear mapping $L_A \in L(L(Y; W); L(X; W))$ or the image of A under the linear mapping $R_B \in L(L(X; Y); L(X; W))$ for which $L_A(B) = R_B(A) = B \circ A$. From this viewpoint $L_A = A^t \otimes I_W$ and $R_B = I_X \otimes B$, where A^t is the dual of A , cf. e.g. Marcus (1973, p. 53). For general f and g , the composition $g \circ f$ can be regarded as the image of g under the linear mapping $L_f \in L(M(Y; W); M(X; W))$ or the image of f under the mapping $R_g \in M(M(X; Y); M(X; W))$ for which $L_f(g) = R_g(f) = g \circ f$. If g is linear then R_g is a linear functional, while L_f is always a linear functional.

We denote by $L_{\otimes} \in L(L(X_1 \otimes \dots \otimes X_r; Y); L(X_1, \dots, X_r; Y))$ the unique linear bijection such that for $\ell \in L(X_1 \otimes \dots \otimes X_r; Y)$,

$$L_{\otimes}(\ell)(x_1, \dots, x_r) = \ell(x_1 \otimes \dots \otimes x_r) = (\ell \circ \otimes^r)(x_1, \dots, x_r)$$

for $x_i \in X_i, i=1, \dots, r$. It is well known that for an arbitrary ϕ belonging to $L(X_1, \dots, X_r; Y)$ there exists a unique mapping ℓ_ϕ belonging to $L(X_1 \otimes \dots \otimes X_r; Y)$, given by $\ell_\phi = L_{\otimes}^{-1}(\phi)$, such that $\ell_\phi \circ \otimes^r = \phi$.

For $X_1 = \dots = X_r = X$ we denote by $L_{\odot} \in L(L(X^{(r)}; Y); L({}^rX; Y))$ the unique linear bijection such that for $\ell \in L(X^{(r)}; Y)$,

$$L_{\odot}(\ell)(x_1, \dots, x_r) = \ell(x_1 \odot \dots \odot x_r) = (\ell \circ \odot^r)(x_1, \dots, x_r)$$

Thus for an arbitrary $\phi \in L({}^rX; Y)$ there exists a unique mapping ℓ_ϕ given by $\ell_\phi = L_{\odot}^{-1}(\phi)|_{X^{(r)}} \in L(X^{(r)}; Y)$, such that $\ell_\phi \circ \odot^r = \phi$, that is,

$$L_{\odot}(\ell_\phi) = \phi. \text{ Hence}$$

$$L_{\odot}^{-1}(\phi) = L_{\otimes}^{-1}(\phi)|_{X^{(r)}}$$

for $\phi \in L({}^rX; Y)$.

3. SOME DIFFERENTIALS ON BANACH SPACES

Let X and Y be Banach spaces over the same field with norms $\|\cdot\|_X$ and $\|\cdot\|_Y$ respectively. For f in $D^q(\mathcal{D}; Y)$, the space of q times differentiable mappings on the non-empty open subset $\mathcal{D}$ of X , the differential of f at $x_0 \in \mathcal{D}$ is defined as the linear mapping, denoted $(Df)(x_0) \in L(X; Y)$, which satisfies

$$(3.1) \quad \|f(x_0 + x_1) - f(x_0) - (Df)(x_0)(x_1)\|_Y = o(\|x_1\|_X)$$

as $x_1 \in X$ tends to 0_X . Here $(Df)(x_0)(x_1) \in Y$, the image of x_1 under $(Df)(x_0)$, is called the derivative of f at x_0 in the direction x_1 or the value at x_1 of the differential of f at x_0 , and is also denoted $f_{x_1}^{(1)}(x_0)$. The differential of f is denoted by $Df \in D^{q-1}(\mathcal{D}; L(X; Y))$, that is, Df is the mapping $x_0 \in \mathcal{D} \rightarrow (Df)(x_0) \in L(X; Y)$. The derivative of f in the direction x_1 is denoted by $f_{x_1}^{(1)} \in D^{q-1}(\mathcal{D}; Y)$.

The r 'th differential of $f \in D^q(\mathcal{D}; Y)$ at x_0 ($r \leq q$) is defined in a recursive manner as the linear mapping $(D^r f)(x_0) \in L(X; L(X; \dots; L(X; Y)))$ defined by

$$(D^r f)(x_0) = (D(D^{r-1} f))(x_0).$$

We denote by $(d^r f)(x_0) \in L({}^r X; Y)$ the image of $(D^r f)(x_0)$ under the natural isometric isomorphism from $L(X; L(X; \dots; L(X; Y)))$ to $L({}^r X; Y)$ by means of

$$(d^r f)(x_0)(x_1, \dots, x_r) = (D^r f)(x_0)(x_1) \dots (x_r).$$

Obviously we have $(df)(x_0) = (Df)(x_0)$. The r 'th derivative of f at x_0 in directions $x_1, \dots, x_r$, that is $(d^r f)(x_0)(x_1, \dots, x_r)$, is also denoted by $f_{x_1, \dots, x_r}^{(r)}(x_0)$, and is the image of x_0 under the r 'th derivative of f in directions $x_1, \dots, x_r$, the mapping of which is denoted $f_{x_1, \dots, x_r}^{(r)} \in D^{q-r}(\mathcal{D}; Y)$. The mapping $(d^r f)(x_0)$ is symmetric in its arguments while the mapping $f_{x_1, \dots, x_r}^{(r)}$ is symmetric in its parameters.

Let $\mathcal{D}$ be an open subset of $X = X_1 \oplus X_2$ (the direct sum of X_1 and

X_2) such that the mapping

$$f_{d_2} : x_1 \in X_1 \rightarrow f(x_1 + d_2) \in Y$$

belongs to $D^q(P_{X_1} \mathcal{D}; Y)$ for every $d_2 \in P_{X_2} \mathcal{D}$, (where P_{X_1} (P_{X_2}) is the projection operator on X_1 (X_2) along X_2 (X_1)); that is, f is q times partially differentiable with respect to X_1 on $\mathcal{D}$. We write $D_{X_1}^q(\mathcal{D}; Y)$ for the space of mappings which are q times partially differentiable with respect to X_1 on $\mathcal{D}$, and for $f \in D_{X_1}^q(\mathcal{D}; Y)$ we write $(D_{X_1}^r f)(x_0)$ for the r 'th ($r \leq q$) partial differential with respect to X_1 at $x_0 \in \mathcal{D}$, and $(d_{X_1}^r f)(x_0)$ for the corresponding isomorphic r -linear mapping. The partial differential of f belonging to $D_{X_1}(\mathcal{D}; Y)$ at $x_0 \in \mathcal{D}$, is the mapping $(D_{X_1} f)(x_0) \in L(X_1; Y)$ such that

$$\| f(x_0 + x_1) - f(x_0) - (D_{X_1} f)(x_0)(x_1) \|_Y = o(\| x_1 \|_X)$$

as $x_1 \in X_1$ tends to 0_{X_1} . If $f \in D(\mathcal{D}; Y)$ then - as is well-known -

$$(Df)(x_0) = (D_{X_1} f)(x_0) \circ P_{X_1} + (D_{X_2} f)(x_0) \circ P_{X_2}$$

so that $(D_{X_1} f)(x_0)$ is then the restriction to X_1 of $(Df)(x_0)$, and similarly for $(D_{X_2} f)(x_0)$.

When $f \in M(X; Y)$ and $g \in M(Y; W)$ belong to $D(\mathcal{D}; Y)$ and $D(f(\mathcal{D}); W)$, respectively, for an open subset $\mathcal{D}$ of X , the chain-rule for the differential of $g \circ f$ at $x_0 \in \mathcal{D}$ is

$$(d(g \circ f))(x_0) = ((dg)(f(x_0))) \circ ((df)(x_0))$$

which can also be written in the alternative forms

$$(d(L_f(g)))(x_0) = L_{(df)(x_0)}((dg)(f(x_0)))$$

and

$$(d(R_g(f)))(x_0) = R_{(dg)(f(x_0))}((df)(x_0)).$$

The following lemma gives examples of some simple and composite differentials. These are well-known results restated in our notation.

LEMMA 3.1.

(i) Let $c_y \in M(X;Y)$, then $c_y \in D^\infty(X;Y)$ and for $x_i \in X$, $i=0,1,\dots,r$

$$(d^r c_y)(x_0)(x_1, \dots, x_r) = \begin{cases} y & \text{if } r=0 \\ 0_Y & \text{if } r \geq 1. \end{cases}$$

(ii) Let $A \in L(X;Y)$, then $A \in D^\infty(X;Y)$ and for $x_i \in X$, $i=0,1,\dots,r$

$$(d^r A)(x_0)(x_1, \dots, x_r) = \begin{cases} Ax_r & \text{if } r=0,1 \\ 0_Y & \text{if } r \geq 2. \end{cases}$$

(iii) Let $A \in L(X;Y)$ and $g \in D^r(Y;W)$, then $L_A(g) \in D^r(X;W)$ and for $x_i \in X$, $i=0,1,\dots,r$,

$$(d^r(L_A(g)))(x_0)(x_1, \dots, x_r) = (d^r g)(Ax_0)(Ax_1, \dots, Ax_r).$$

(iv) Let $f \in D^r(X;Y)$ and $B \in L(Y;W)$, then $R_B(f) \in D^r(X;W)$ and for $x_i \in X$, $i=0,1,\dots,r$,

$$(d^r(R_B(f)))(x_0)(x_1, \dots, x_r) = B(d^r f)(x_0)(x_1, \dots, x_r).$$

PROOF. (i): Since $(d^0 c_y)(x_0) = c_y(x_0) = y$, the result for $r=0$ is proved.

Further $\|c_y(x_0 + x_1) - c_y(x_0)\|_Y = \|0_Y\|_Y = o(\|x_1\|_X)$, which proves that $(c_y)_{x_1}^{(1)} = c_{0_Y} \in M(X;Y)$ and the result for $r=1$. Hence by repeated use of

$$(c_y)_{x_1, \dots, x_r}^{(r)} = (c_{0_Y})_{x_2, \dots, x_r}^{(r-1)}$$

the result for $r \geq 2$ is obtained.

(ii): The result for $r=0$ is a rewriting of Ax_0 as $(d^0 A)(x_0)$. Further

$\|A(x_0 + x_1) - Ax_0 - Ax_1\|_Y = \|0_Y\|_Y = o(\|x_1\|_X)$, which proves that

$A_{x_1}^{(1)} = c_{Ax_1} \in M(X;Y)$ and the result for $r=1$. The result for $r \geq 2$ follows

by using that $A_{x_1, \dots, x_r}^{(r)} = (c_{Ax_1})_{x_2, \dots, x_r}^{(r-1)}$ and the result in (i).

(iii): By the chain-rule we have

$$(d(L_A(g)))(x_0)(x_1) = L_{(dA)(x_0)}((dg)(Ax_0))(x_1) = (dg)(Ax_0)(Ax_1)$$

where, in the second step, we have used that $(dA)(x_0) = A$ and that $L_A(h)(x_1) = h(A(x_1))$, by definition. Hence we have proved the result for $r=1$ and that $(L_A(g))_{x_1}^{(1)} = L_{A_{Ax_1}}(g_{Ax_1})^{(1)}$. The result for $r \geq 2$ is then obtained by using

$$(L_A(g))_{x_1, \dots, x_r}^{(r)} = (L_{A_{Ax_1}}(g_{Ax_1}))_{x_2, \dots, x_r}^{(r-1)}$$

recursively.

(iv): By the chain-rule we have

$$(d(R_B(f)))(x_0)(x_1) = R_{(dB)(f(x_0))}((df)(x_0))(x_1) = B((df)(x_0)(x_1))$$

and hence the result is true for $r=1$ and $(R_B(f))_{x_1}^{(1)} = R_B(f_{x_1})^{(1)}$. By using

$$(R_B(f))_{x_1, \dots, x_r}^{(r)} = (R_B(f_{x_1}))_{x_2, \dots, x_r}^{(r-1)}$$

recursively, the result for $r \geq 2$ is obtained. $\square$

The results of Lemma 3.1 are stated in terms of derivatives at a certain point and in certain directions. It is sometimes convenient to have the results stated in terms of mappings, that is, in terms of the differentials at the specific point. It should not cause any problem to obtain expressions for these mappings, except for the result in (iii). In this case it is more convenient to write the result in terms of the corresponding mapping defined on the completely symmetric space:

$$(3.2) \quad L_{\odot}^{-1}((d^r(L_A(g)))(x_0)) = L_{\odot}^{-1}((d^r g)(Ax_0)) \circ A^{<r>}.$$

When calculating differentials in coordinate spaces, which is convenient for computation, it is sometimes - as we shall see in section 5 - easier to calculate a differential in a larger space X' than in the space X . Let $X' = X \oplus W$ and let $f \in D^r(\mathcal{D}; Y)$ where $\mathcal{D}$ is an open subset of X . Denote by f' the extended mapping on X' defined by

$$f'(x') = f(P_X x') = L_{P_X}(f)(x')$$

for $x' \in X'$ and where P_X is the projection on X along W . Then

$f' \in D^r(\mathcal{D}'; \mathcal{V})$ where $\mathcal{D}' = \mathcal{D} \oplus \mathcal{W}$, and according to (3.2),

$$(3.3) \quad L_{\odot}^{-1}((d^r f')(x'_0)) = L_{\odot}^{-1}((d^r f)(P_X x'_0)) \circ P_X^{<r>}$$

for $x'_0 \in \mathcal{D}'$. Also, from the construction, f' is constant along $\mathcal{W}$ and thus

$(D_{\mathcal{W}} f')(x'_0) = c_{0_{\mathcal{V}}} \in M(X'; \mathcal{V})$ for $x'_0 \in \mathcal{D}'$ according to Lemma 3.1 (i). Hence

$$(3.4) \quad (D_X f')(x'_0) \circ P_X = (Df)(P_X x'_0) \circ P_X$$

or, in other words, $(D_X f')(x'_0)$ and $(Df)(P_X x'_0)$ coincide on X .

The following theorem concerns differentials of power mappings on X .

THEOREM 3.1. Let $\pi_X^k \in M(X; X^{<k>})$, then $\pi_X^k \in D^{\infty}(X; X^{<k>})$ and for $x_i \in X$, $i=0, 1, \dots, r$,

$$\begin{aligned} & (d^r \pi_X^k)(x_0)(x_1, \dots, x_r) \\ &= \begin{cases} \binom{k}{r, k-r} S_{r, k-r}(x_1 \otimes \dots \otimes x_r \otimes x_0^{<k-r>}) & r \leq k \\ 0_{X^{<k>}} & r > k. \end{cases} \quad \square \end{aligned}$$

To prove this generalization of derivatives of monomials on K , we show the slightly more general result given in the following lemma.

LEMMA 3.2. Denote by $\sum_j^m$ the summation over the set of all non-negative integers $u_{1,j}, \dots, u_{m,j}$ such that $u_{1,j} + \dots + u_{m,j} = 1$. Let $\underline{u}_{k,r} = (u_{k,1}, \dots, u_{k,r})$ and $|\underline{u}_{k,r}| = u_{k,1} + \dots + u_{k,r}$, $k=1, \dots, m$. Further let $c_{y_k} \in M(X; \mathcal{V}_k)$, $k=1, \dots, m$, and $i = i_1 + \dots + i_m$. Then it holds that

$$\begin{aligned} & (d^r (\bigotimes_{k=1}^m (c_{y_k} \otimes I_X^{<i_k>})) \circ \pi_X^{m+i})(x_0)(x_1, \dots, x_r) \\ &= \begin{cases} \sum_1^m \dots \sum_r^m \left(\prod_{k=1}^m \binom{i_k}{\underline{u}_{k,r}} \right) \times \\ \times \left(\bigotimes_{k=1}^m (y_k \otimes S_{\underline{u}_{k,r}, i_k - |\underline{u}_{k,r}|}((\bigotimes_{j=1}^r x_j^{<u_{k,j}>}) \otimes x_0^{<i_k - |\underline{u}_{k,r}|>})) \right) & r \leq i \\ 0_{\bigotimes_{k=1}^m (\mathcal{V}_k \otimes X^{<i_k>})} & r > i. \end{cases} \end{aligned}$$

PROOF. First we prove the result for $r=1$. By aid of (2.6) for $k=2$, we obtain

$$\begin{aligned}
 & \left\| \left(\bigotimes_{k=1}^m (c_{y_k} \otimes I_X^{\langle i_k \rangle}) \circ \pi_X^{m+i} \right) (x_0 + x_1) \right. \\
 & \quad \left. - \left(\bigotimes_{k=1}^m (c_{y_k} \otimes I_X^{\langle i_k \rangle}) \circ \pi_X^{m+i} \right) (x_0) \right. \\
 & \quad \left. - \sum_{\substack{u_1=0 \\ u_1+\dots+u_m=1}}^{i_1} \dots \sum_{u_m=0}^{i_m} \left(\prod_{k=1}^m \binom{i_k}{u_k} \right) \right. \\
 & \quad \left. \left(\bigotimes_{k=1}^m (y_k \otimes S_{u_k, i_k-u_k}^{\langle u_k \rangle} (x_1^{\langle u_k \rangle} \otimes x_0^{\langle i_k-u_k \rangle})) \right) \right\| \bigotimes_{k=1}^m (y_k \otimes X^{\langle i_k \rangle}) \\
 & = \left\| \sum_{\substack{u_1=0 \\ u_1+\dots+u_m > 1}}^{i_1} \dots \sum_{u_m=0}^{i_m} \left(\prod_{k=1}^m \binom{i_k}{u_k} \right) \right. \\
 & \quad \left. \left(\bigotimes_{k=1}^m (y_k \otimes S_{u_k, i_k-u_k}^{\langle u_k \rangle} (x_1^{\langle u_k \rangle} \otimes x_0^{\langle i_k-u_k \rangle})) \right) \right\| \bigotimes_{k=1}^m (y_k \otimes X^{\langle i_k \rangle}) \\
 & \leq \sum_{\substack{u_1=0 \\ u_1+\dots+u_m > 1}}^{i_1} \dots \sum_{u_m=0}^{i_m} \prod_{k=1}^m \binom{i_k}{u_k} \|y_k\|_{Y_k} \|S_{u_k, i_k-u_k}\| \|x_1\|_X^u \|x_0\|_X^{i-u}
 \end{aligned}$$

where $u = u_1 + \dots + u_m$. This expression is of order $o(\|x_1\|_X)$ which shows that

$$\begin{aligned}
 & \left(d \left(\bigotimes_{k=1}^m (c_{y_k} \otimes I_X^{\langle i_k \rangle}) \circ \pi_X^{m+i} \right) (x_0) (x_1) \right. \\
 & \quad \left. - \sum_{\substack{u_1=0 \\ u_1+\dots+u_m=1}}^{i_1} \dots \sum_{u_m=0}^{i_m} \left(\prod_{k=1}^m \binom{i_k}{u_k} \right) \right. \\
 & \quad \left. \left(\bigotimes_{k=1}^m (y_k \otimes S_{u_k, i_k-u_k}^{\langle u_k \rangle} (x_1^{\langle u_k \rangle} \otimes x_0^{\langle i_k-u_k \rangle})) \right) \right)
 \end{aligned}$$

and hence we have proved the result for $r=1$.

Next, assume that the result is true for $r=s < i$. By Lemma 3.1 (iv) we then obtain

$$\begin{aligned}
 & \left(\bigotimes_{k=1}^m (c_{y_k} \otimes I_X^{\langle i_k \rangle}) \right) \circ \pi_X^{m+1}(s+1)_{x_1, \dots, x_{s+1}} \\
 &= \sum_1^m \dots \sum_s^m \left(\prod_{k=1}^m \left[\begin{matrix} i_k \\ u_{k,s} \end{matrix} \right]_+ \right) \left(\bigotimes_{k=1}^m (I_{Y_k} \otimes S_{u_{k,s}, i_k - |u_{k,s}|}) \right) \\
 & \quad \left(\bigotimes_{k=1}^m (c_{y_k} \otimes \left(\bigotimes_{j=1}^s c_{x_j}^{\langle u_{k,j} \rangle} \right) \otimes I_X^{\langle i_k - |u_{k,s}| \rangle} \right)_{x_{s+1}}^{(1)} \\
 &= \sum_1^m \dots \sum_s^m \left(\prod_{k=1}^m \left[\begin{matrix} i_k \\ u_{k,s} \end{matrix} \right]_+ \right) \left(\bigotimes_{k=1}^m (I_{Y_k} \otimes S_{u_{k,s}, i_k - |u_{k,s}|}) \right) \\
 & \quad \sum_{\substack{u_{1,s+1}=0 \\ u_{1,s+1} + \dots + u_{m,s+1}=1}}^{i_1 - |u_{1,s}|} \dots \sum_{\substack{u_{m,s+1}=0 \\ u_{1,s+1} + \dots + u_{m,s+1}=1}}^{i_m - |u_{m,s}|} \left(\prod_{k=1}^m \left[\begin{matrix} i_k - |u_{k,s}| \\ u_{k,s+1} \end{matrix} \right]_+ \right) \\
 & \quad \left(\bigotimes_{k=1}^m (I_{Y_k} \otimes I_X^{\langle |u_{k,s}| \rangle} \otimes S_{u_{k,s+1}, i_k - |u_{k,s}| - u_{k,s+1}}) \right) \\
 & \quad \left(\bigotimes_{k=1}^m (y_k \otimes \left(\bigotimes_{j=1}^s x_j^{\langle u_{k,j} \rangle} \right) \otimes x_{s+1}^{\langle u_{k,s+1} \rangle} \otimes x_0^{\langle i_k - |u_{k,s}| - u_{k,s+1} \rangle} \right) \\
 &= \sum_1^m \dots \sum_{s+1}^m \left(\prod_{k=1}^m \left[\begin{matrix} i_k \\ u_{k,s+1} \end{matrix} \right]_+ \right) \\
 & \quad \left(\bigotimes_{k=1}^m (y_k \otimes S_{u_{k,s+1}, i_k - |u_{k,s+1}|} \left(\bigotimes_{j=1}^{s+1} x_j^{\langle u_{k,j} \rangle} \right) \otimes x_0^{\langle i_k - |u_{k,s+1}| \rangle} \right)
 \end{aligned}$$

where the last step follows by aid of Lemma 2.1. The result is thus also true for $r=s+1 \leq i$. An induction argument now shows that the result is true for $r=1, \dots, i$, since it is true for $r=1$ as shown above.

Further

$$\begin{aligned}
 & \left(\bigotimes_{k=1}^m (c_{y_k} \otimes I_X^{\langle i_k \rangle}) \right) \circ \pi_X^{m+i}(i)_{x_1, \dots, x_i} \\
 &= \sum_1^m \dots \sum_i^m \left(\prod_{k=1}^m \left[\begin{matrix} i_k \\ u_{k,i} \end{matrix} \right]_+ \right) \left(\bigotimes_{k=1}^m (I_{Y_k} \otimes S_{u_{k,i}, i_k - |u_{k,i}|}) \right) \\
 & \quad \circ \left(\bigotimes_{k=1}^m (c_{y_k} \otimes \left(\bigotimes_{j=1}^i c_{x_j}^{\langle u_{k,j} \rangle} \right) \circ \pi_X^{\langle i \rangle} \right)
 \end{aligned}$$

which is a constant map, and hence the result for $r > i$ immediately follows from Lemma 3.1 (i). □

PROOF OF THEOREM 3.1. The result immediately follows from Lemma 3.2 by taking $m = 1$, $i_1 = k$, $Y_1 = K$ and $y_1 = 1_K$. $\square$

By combining Lemma 3.1 (iv) and Theorem 3.1, we obtain the following useful result.

COROLLARY 3.1. Let $\pi_X^k \in M(X; X^{<k>})$ and let $\ell \in L(X^{<k>}; Y)$, then $\ell \circ \pi_X^k \in D^\infty(X; Y)$ and for $x_i \in X$, $i=0,1,\dots,r$,

$$\begin{aligned} & (d^r(\ell \circ \pi_X^k))(x_0)(x_1, \dots, x_r) \\ &= \begin{cases} \left[\begin{smallmatrix} k \\ \underline{1}_r, k-r \end{smallmatrix} \right] \ell(S_{\underline{1}_r, k-r}(x_1 \times \dots \times x_r \times x_0^{<k-r>})) & r \leq k \\ 0_Y & r > k. \end{cases} \quad \square \end{aligned}$$

Especially we may take ℓ in Corollary 3.1 to be the adjoint of a decomposable element in $X^{<k>}$. If this element is taken to be the image of $x \in X$ under the tensor power mapping π_X^k , it is clear that $x^{<k>*} \circ S_{\underline{1}_r, k-r} = (S_{\underline{1}_r, k-r}^*(x^{<k>}))^* = x^{<k>*}$ and thus the following result holds.

COROLLARY 3.2. Let X be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$, then for $x \in X$, $x^{<k>*} \circ \pi_X^k \in D^\infty(X; K)$ and for $x_i \in X$, $i=0,1,\dots,r$,

$$\begin{aligned} & (d^r(x^{<k>*} \circ \pi_X^k))(x_0)(x_1, \dots, x_r) \\ &= \begin{cases} \left[\begin{smallmatrix} k \\ \underline{1}_r, k-r \end{smallmatrix} \right] \left(\prod_{i=1}^r \langle x_i, x \rangle \right) \langle x_0, x \rangle^{k-r} & r \leq k \\ 0_K & r > k. \end{cases} \quad \square \end{aligned}$$

By aid of this result and the exponential expansion

$$(3.5) \quad \exp \circ f = \sum_{k=0}^{\infty} \pi_K^k \circ f/k!$$

for $f \in M(X; K)$, we obtain the result on which the next section is based.

COROLLARY 3.3. Let X be a Hilbert space over K with inner product

$\langle \cdot, \cdot \rangle$. Let $x, x_i \in X, i=0,1,\dots,r$; then $\exp \circ x^* \in D^\infty(X;K)$ and

$$(d^r(\exp \circ x^*))(x_0)(x_1, \dots, x_r) = \exp(\langle x_0, x \rangle) \prod_{i=1}^r \langle x_i, x \rangle.$$

PROOF. Since $\pi_K^k \circ x^* = x^{\langle k \rangle^*} \circ \pi_X^k$ we have by (3.5) and Corollary 3.2

$$\begin{aligned} & (d^r(\exp \circ x^*))(x_0)(x_1, \dots, x_r) \\ &= \sum_{k=0}^{\infty} (d^r(x^{\langle k \rangle^*} \circ \pi_X^k))(x_0)(x_1, \dots, x_r)/k! \\ &= \sum_{k=r}^{\infty} \left(\prod_{i=1}^r \langle x_i, x \rangle \right) \langle x_0, x \rangle^{k-r} / (k-r)! = \left(\prod_{i=1}^r \langle x_i, x \rangle \right) \exp(\langle x_0, x \rangle). \quad \square \end{aligned}$$

4. MOMENTS AND CUMULANTS FROM DIFFERENTIALS OF GENERATING FUNCTIONS

4.1. Introduction

Moments for a multivariate random variable are usually defined as scalars being the expectation of products of the coordinates in a coordinate space representation of the random variable. As in the case of the mean and dispersion of a random vector, these scalars can be arranged in an array which then represents the mean vector and dispersion matrix of the random vector.

Cumulants for a multivariate random variable are usually defined as scalars being the joint cumulants of the coordinates in a coordinate space representation of the random variable. These joint cumulants are the coefficients in the Taylor's series expansion of the logarithm of the joint moment generating function of the coordinates; cf. e.g. Brillinger (1975, Sec 2.3) and Speed (1983). An arrangement of the scalars as in the case of the mean and dispersion matrix is not customary.

In this section we shall define moments and cumulants of multivariate random variables as elements in the completely symmetric space generated by the space in which the random variable take values. We obtain these elements by differentiating the generating functions for the statistics.

4.2. Moments and cumulants of Hilbert space-valued random variables

Let X be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle_X$. Denote by $(\Omega, \mathcal{B}(\Omega), P)$ some basic probability space and consider a weakly measurable mapping $x: \Omega \rightarrow X$, i.e., $f(x)$ is measurable with respect to $\mathcal{B}(\Omega)$, the σ -algebra of subsets of Ω , and the Borel σ -algebra on the real line, for all $f \in X^*$. We denote by $\mathcal{A}_X$ the smallest of all σ -algebras that contain sets of the form

$$\{y; f(y) < t\}, f \in X^*, -\infty < t < \infty,$$

for all $f \in X^*$ and all real t . For a separable space X , this σ -algebra coincides with $\mathcal{B}_X$, the σ -algebra of Borel sets (with respect to the norm $\|\cdot\|_X$) of X , cf. Vakhania (1981, Sec. 4.1). We denote by ν_x the probability measure on $(X, \mathcal{A}_X)$ induced by x and given by

$$\nu_x(A) = P(\{\omega; x(\omega) \in A\}), \quad A \in \mathcal{A}_X.$$

The symbol $E(x)$ stands for the expectation of x in the Pettis integral sense (Pettis 1938, Def. 2.1), that is, $E(x)$ (if it exists) is the unique element in X that satisfies

$$f(E(x)) = \int_X f(y) \nu_x(dy) = \int_{\Omega} f(x(\omega)) P(d\omega)$$

for all $f \in X^*$ (Mourier 1953, p. 164) and which is determined by the set of values $\{f(E(x)); f \in X^*\}$ (Hille 1948, p. 22). For brevity we write

$$E(x) = \int_X y \nu_x(dy)$$

for this element.

The moment generating function (m.g.f.) M_x and the cumulant generating function (c.g.f.) C_x , of the random variable x are defined as

$$(4.1) \quad M_x = E(\exp \circ x^*) = \int_X \exp \circ y^* \nu_x(dy) \in M(T; K)$$

and

$$(4.2) \quad C_x = \ln \circ M_x \in M(T; K)$$

where $\ln$ is the natural logarithm. The set of definition is the subset T of X for which

$$M_x(t) = \int_X \exp(\langle t, y \rangle_X) \nu_x(dy)$$

is positive and finite for all $t \in T$.

Firstly, we assume that M_x and thus also C_x belong to $D^\infty(\mathcal{D}; K)$ where $\mathcal{D}$ is an open subset of T containing 0_X . Moment and cumulant mappings are in this case defined as differentials at zero of the m.g.f.

and c.g.f. respectively; the r 'th moment about zero mapping of x is given by

$$(4.3) \quad \hat{\mu}_{r,x} = (d^r M_x)(0_X) \quad r = 0, 1, \dots$$

and the r 'th cumulant mapping of x is given by

$$(4.4) \quad \hat{\kappa}_{r,x} = (d^r C_x)(0_X) \quad r = 0, 1, \dots$$

The r -linear mappings $\hat{\mu}_{r,x}$ and $\hat{\kappa}_{r,x}$ belong to $L_{\odot}({}^r T; K)$, but by using the multilinear property, these mappings can naturally be extended to $L_{\odot}({}^r \text{lc}(T); K)$. Since $T \subset \mathcal{D}$ which is open and contains 0_X , it follows that $\text{lc}(T) = X$. These augmented versions in $L_{\odot}({}^r X; K)$ will be considered in the following, and for brevity in notation we shall let $\hat{\mu}_{r,x}$ and $\hat{\kappa}_{r,x}$ denote the extended mappings.

Moment (cumulant) adjoints are defined as the images of the moment (cumulant) mappings under the linear mappings which relates a specific symmetric multilinear mapping to the corresponding linear mapping on the completely symmetric space. Hence the r 'th moment about zero adjoint of x is given by

$$(4.5) \quad \mu_{r,x}^* = L_{\odot}^{-1}(\hat{\mu}_{r,x}) \in L(X^{(r)}; K) = X^{(r)*} \quad r = 0, 1, \dots$$

and the r 'th cumulant adjoint of x is given by

$$(4.6) \quad \kappa_{r,x}^* = L_{\odot}^{-1}(\hat{\kappa}_{r,x}) \in L(X^{(r)}; K) = X^{(r)*} \quad r = 0, 1, \dots$$

Since $\mu_{r,x}^*$ and $\kappa_{r,x}^*$ are linear mappings, they can be considered as adjoints - as is indicated by their name - of unique elements, $\mu_{r,x}$ and $\kappa_{r,x}$ respectively, in $X^{(r)}$ such that $\mu_{r,x}^*(y) = \langle y, \mu_{r,x} \rangle_{X^{(r)}} / k^r$ and $\kappa_{r,x}^*(y) = \langle y, \kappa_{r,x} \rangle_{X^{(r)}} / k^r$ for all $y \in X^{(r)}$ and where $k = \langle 1_K, 1_K \rangle_K$. The elements $\mu_{r,x}$ and $\kappa_{r,x}$ are called the r 'th moment about zero and the r 'th cumulant, respectively, of x . By interchanging the order of differentiation and integration, we obtain by aid of Corollary 3.3

$$\hat{\mu}_{r,x} = (d^r (\int_X \exp \circ y^* v_x(dy))) (0_X) = \int_X (d^r (\exp \circ y^*)) (0_X) v_x(dy)$$

$$\begin{aligned}
 &= \int_X k^r y^{<r>*} \circ \odot^r v_x(dy) = E(k^r x^{<r>*} \circ \odot^r) = k^r E(x^{<r>*}) \circ \odot^r \\
 &= L_{\odot}(k^r E(x^{<r>*})).
 \end{aligned}$$

Hence

$$\mu_{r,x}^* = k^r E(x^{<r>*}) = k^r E(x^{<r>})^* \in L(X^{(r)}; K) = X^{<r>*}.$$

and thus

$$(4.7) \quad \mu_{r,x} = E(x^{<r>}).$$

The definition of the r 'th moment through the m.g.f. thus yields an expression which is a natural extension of $E(x^r)$ for scalar random variables. We may further write

$$(4.8) \quad M_x = \sum_{i=0}^{\infty} \mu_{i,x}^* \circ \pi_X^i / i! \quad (\mu_{0,x}^* = c_{1_K})$$

since then by Corollary 3.1

$$(d^r M_x)(0_X) = L_{\odot}(\mu_{r,x}^*)$$

which is the relation (4.5) in another form. Similarly the c.g.f. can be written

$$(4.9) \quad C_x = \sum_{i=0}^{\infty} \kappa_{i,x}^* \circ \pi_X^i / i! \quad (\kappa_{0,x}^* = c_{0_K})$$

since then by Corollary 3.1 we obtain

$$(d^r C_x)(0_X) = L_{\odot}(\kappa_{r,x}^*)$$

which is the relation (4.6).

The series in (4.8) and (4.9) are formal, indicating equality for the image of $t \in T$ being such that the series converge.

Remark 4.1. The difference between $\mu_{r,x}$ and $\kappa_{r,x}$ on one hand and $\mu_{r,x}^*$ and $\kappa_{r,x}^*$ on the other is that the former are elements in a space and the latter are operators on a space. The approach used in literature is not always consistent; the mean of a random vector is often seen as an element in the space of the random element while the covariance is

conceived as an operator. However the operator is dependent on the inner product on K , which may be a disadvantage. $\square$

The relations (4.1), (4.2), (4.8) and (4.9) yield

$$(4.10) \quad \exp \circ \left(\sum_{i=0}^{\infty} \kappa_{i,x}^* \circ \pi_X^i / i! \right) = \sum_{i=0}^{\infty} \mu_{i,x}^* \circ \pi_X^i / i!$$

which is an analogue of a definition of cumulants given by Thiele (1899) (although not the original one). Cf. also Hald (1981).

4.3. Relations between moments and cumulants

The cumulant of a certain order, defined by the cumulant generating function, can as in the scalar case be related to moments of the same and lower order. We distinguish between three types of relations: Non-recursive relations, recursive relations and differential relations.

(a) Non-recursive relations

The results of the next theorem relate moments about zero to cumulants and vice versa and are generalizations to completely symmetric spaces of the relations found by Thiele (1899) for scalar random variables.

THEOREM 4.1. Let $\underline{j}_r = (j_1, \dots, j_r)$ be a partition of weight r of j and $\underline{j}_m^+ = (j_1^+, \dots, j_m^+)$ a positive partition of weight m of j . Let $\sum_{\underline{j}_r}^r$ denote the summation over the set of partitions of weight r such that $j_1 + 2j_2 + \dots + rj_r = r$ and let $\sum_{\underline{i}_m, \underline{j}_m^+}^r$ denote the summation over the set of $(\underline{i}_m, \underline{j}_m^+)$, $\underline{i}_m = (i_1, \dots, i_m)$ such that $i_1 j_1^+ + i_2 j_2^+ + \dots + i_m j_m^+ = r$, for integers $0 < i_1 < \dots < i_m$. Let $C_{\underline{j}_r} = r! / \prod_{i=1}^r (j_i! i!^{j_i})$ and $C_{\underline{j}_m^+}^+ = r! / \prod_{k=1}^m (j_k^+! i_k!^{j_k^+})$; then the following results hold for $r=1, 2, \dots$.

$$(i) \quad \mu_{r,x} = \sum_{j=1}^r \sum_{\underline{j}_r}^r C_{\underline{j}_r} S_{\underline{j}_r} \left(\bigotimes_{i=1}^r \kappa_{i,x}^{<j_i>} \right)$$

$$(ii) \quad \mu_{r,x} = \sum_{m=1}^r \sum_{\underline{j}_m, \underline{j}_m}^r C_{\underline{j}_m}^+ S_{\underline{j}_m} \left(\bigotimes_{k=1}^m \kappa_{i_k, x}^{<j_k^+>} \right)$$

$$(iii) \quad \kappa_{r,x} = \sum_{j=1}^r \sum_{\underline{j}_r}^r (-1)^{j-1} (j-1)! C_{\underline{j}_r} S_{\underline{j}_r} \left(\bigotimes_{i=1}^r \mu_{i,x}^{<j_i>} \right)$$

$$(iv) \quad \kappa_{r,x} = \sum_{m=1}^r \sum_{\underline{j}_m, \underline{j}_m}^r (-1)^{j-1} (j-1)! C_{\underline{j}_m}^+ S_{\underline{j}_m} \left(\bigotimes_{k=1}^m \mu_{i_k, x}^{<j_k^+>} \right)$$

PROOF. (i): By exponential series expansion and by aid of (2.3) we find that

$$\begin{aligned} \exp\left(\sum_{i=1}^{\infty} \langle t^{<i>}, \kappa_{i,x} \rangle / i!\right) &= \prod_{i=1}^{\infty} \exp(\langle t^{<i>}, \kappa_{i,x} \rangle / i!) \\ &= \prod_{i=1}^{\infty} \left(\sum_{j=0}^{\infty} (\langle t^{<i>}, \kappa_{i,x} \rangle / i!)^j / j! \right) = \prod_{i=1}^{\infty} \left(\sum_{j=0}^{\infty} \langle t^{<ij>}, \kappa_{i,x}^{<j>} \rangle / (j! i!^j) \right) \\ &= \sum' \prod_{i=1}^{\infty} (\langle t^{<ij>}, \kappa_{i,x}^{<j>} \rangle / (j! i!^j)) \\ (4.11) \quad &= \sum' \langle t^{\langle \Sigma'' ij_i \rangle}, \bigotimes_{i=1}^{\infty} \kappa_{i,x}^{<j_i>} \rangle / \left(\prod_{i=1}^{\infty} (j_i! i!^{j_i}) \right) \end{aligned}$$

Here the summation $\sum'$ is over all $j_i=0,1,\dots$, for $i=1,2,\dots$, and Σ'' is the sum over all $i=1,2,\dots$. By (4.2) and (4.9) we conclude that the first expression is equal to $M_x(t)$. Hence by taking the r th differential at $t=0$ of (4.11), we obtain $\hat{\mu}_{r,x}$. Thus by (2.1) and Corollary 3.1

$$\hat{\mu}_{r,x} = \sum'_{\Sigma'' ij_i=r} L_{\odot} \left(k(r! / \left(\prod_{i=1}^{\infty} (j_i! i!^{j_i}) \right)) \left(\bigotimes_{i=1}^{\infty} \kappa_{i,x}^{<j_i>} \right)^* \circ S_{\underline{j}_r} \right)$$

where $k = \langle 1_K, 1_K \rangle_K$. In fact here the tensor product and the product are both finite since the restriction $\Sigma'' ij_i=r$ implies that $j_i=0$ for $i>r$. The restriction further implies that $j_i \leq r$ for $i \leq r$ and thus the summation is performed over the same set as $\sum_{j=1}^r \sum_{\underline{j}_r}^r$. By (4.5) we thus obtain

$$\mu_{r,x}^* = \sum_{j=1}^r \sum_{\underline{j}_r}^r k(r! / \prod_{i=1}^r j_i! i!^{j_i}) \left(\bigotimes_{i=1}^r \kappa_{i,x}^{<j_i>} \right)^* \circ S_{\underline{j}_r}$$

and hence

$$\begin{aligned}
 \mu_{r,x}^*(y) &= \sum_{j=1}^r \sum_{\underline{j}_r}^r k(r! / \prod_{i=1}^r j_i! i!^{j_i}) (\bigotimes_{i=1}^r \kappa_{i,x}^{<j_i>})^* (S_{\underline{1}_r}(y)) \\
 &= \sum_{j=1}^r \sum_{\underline{j}_r}^r (r! / \prod_{i=1}^r j_i! i!^{j_i}) \langle S_{\underline{1}_r}(y), \bigotimes_{i=1}^r \kappa_{i,x}^{<j_i>} \rangle_{\chi^{<r>}} \\
 &= \sum_{j=1}^r \sum_{\underline{j}_r}^r (r! / \prod_{i=1}^r j_i! i!^{j_i}) \langle y, S_{\underline{1}_r}(\bigotimes_{i=1}^r \kappa_{i,x}^{<j_i>}) \rangle_{\chi^{<r>}}
 \end{aligned}$$

where in the last step we have also used Theorem 2.1 (iv). Since this is true for all $y \in \chi^{(r)}$, the result follows.

(ii): By dividing the sum of (i) into two parts, one summing over the number of positive summation variables of $j_1, \dots, j_r$ and the second summing over the values of these, the result follows.

(iii): By logarithmic series expansion and by aid of (2.2) we obtain

$$\begin{aligned}
 \ln(\sum_{i=0}^{\infty} \langle t^{<i>} \rangle_{\mu_{i,x}} / i!) &= \sum_{j=1}^{\infty} (-1)^{j-1} (\sum_{i=1}^{\infty} \langle t^{<i>} \rangle_{\mu_{i,x}} / i!)^j / j \\
 &= \sum_{j=1}^{\infty} (-1)^{j-1} (\sum'_{\Sigma'' j_i=j} j! \prod_{i=1}^{\infty} \langle t^{<i>} \rangle_{\mu_{i,x}}^{j_i} / \prod_{i=1}^{\infty} (j_i! i!^{j_i})) / j \\
 &= \sum_{j=1}^{\infty} (-1)^{j-1} (j-1)! (\sum'_{\Sigma'' j_i=j} \prod_{i=1}^{\infty} \langle t^{<ij_i>} \rangle_{\mu_{i,x}}^{<j_i>} / \prod_{i=1}^{\infty} (j_i! i!^{j_i})) \\
 &= \sum_{j=1}^{\infty} (-1)^{j-1} (j-1)! (\sum'_{\Sigma'' j_i=j} \langle t^{\Sigma'' ij_i} \rangle_{\mu_{i,x}}^{<j_i>} / \prod_{i=1}^{\infty} (j_i! i!^{j_i})).
 \end{aligned}$$

By a similar argument as in the proof of (i) we thus obtain

$$\hat{\kappa}_{r,x} = \sum_{j=1}^{\infty} (-1)^{j-1} (j-1)! (\sum'_{\Sigma'' j_i=j} L_{\odot} (k(r! / \prod_{i=1}^{\infty} (j_i! i!^{j_i})) (\bigotimes_{i=1}^{\infty} \mu_{i,x}^{<j_i>})^* \circ S_{\underline{1}_r}))$$

$\Sigma'' ij_i=r$

by aid of (2.1) and Corollary 3.1. Because of the restriction $\Sigma'' ij_i=r$, this can be written

$$\hat{\kappa}_{r,x} = \sum_{j=1}^r (-1)^{j-1} (j-1)! \sum_{\underline{j}_r}^r L_{\odot} (k(r! / \prod_{i=1}^r (j_i! i!^{j_i})) (\bigotimes_{i=1}^r \mu_{i,x}^{<j_i>})^* \circ S_{\underline{1}_r}).$$

By a similar argument as in the proof of (i), the result follows.

(iv): This result is obtained by a similar argument as in the proof of (ii). $\square$

Remark 4.2. Let $\gamma_i \in X^{(1)}$, $i=1, \dots, r$ and $\sum j_i = r$. It is not generally true that $\bigotimes_{i=1}^r \gamma_i^{<j_i>} \in X^{(r)}$. However $\bigotimes_{i=1}^r \gamma_i^{<j_i>} \in X^{<r>}$ and the projection $S_{\frac{1}{-r}} \left(\bigotimes_{i=1}^r \gamma_i^{<j_i>} \right)$ of $\bigotimes_{i=1}^r \gamma_i^{<j_i>}$ belongs to $X^{(r)}$ since $S_{\frac{1}{-r}}$ is the projection operator of $X^{<r>}$ on $X^{(r)}$ along $X^{(r)\perp}$ (Theorem 2.1 (iii) and (iv)). Hence it follows that the right-hand sides of the expressions in Theorem 4.1 (i)-(iv) all belong to $X^{(r)}$. $\square$

Remark 4.3. For vector spaces of dimension 1, we have that

$$S_{\frac{1}{-r}} \left(\bigotimes_{k=1}^m \gamma_{i_k, x}^{<j_k>} \right) = \prod_{k=1}^m \gamma_{i_k, x}^{j_k}$$

for $\gamma_{i_k, x}$ being $\mu_{i_k, x}$ or $\kappa_{i_k, x}$, and hence the expressions of Theorem 4.1 (ii) and (iv) reduces in this case to the expressions of Thiele (1899), cf. also Hald (1981, p. 8) or expressions (3.33) and (3.39) of Kendall and Stuart (1976) (noting the misprint "nonnegative" immediately below (3.33) which should read "positive"). $\square$

Remark 4.4. The order in which the moments and cumulants appear on the right hand sides of the expressions in Theorem 4.1 (i)-(iv) is immaterial. For example this is seen in (i) by writing an arbitrary order as

$$P_r^{\otimes}(\xi) \left(\bigotimes_{i=1}^r \kappa_{i, x}^{<j_i>} \right)$$

for some $\xi \in S_{\frac{1}{-r}}$. But according to Theorem 2.1 (i), $S_{\frac{1}{-r}} \circ P_r^{\otimes}(\xi) = S_{\frac{1}{-r}}$ and hence the same result is obtained. $\square$

(b) Recursive relations

The original definition by Thiele (1889, p. 19) of cumulants was given in terms of a recursive formula:

$$\mu_{r+1, x} = \sum_{i=0}^r \binom{r}{i}_+ \mu_{r-i, x} \kappa_{i+1, x}$$

(although Thiele considered sample versions of moments and cumulants) or

$$\mu_{r,x} = \sum_{i=1}^r \binom{r-1}{i-1}_+ \mu_{r-i,x} \kappa_{i,x}$$

cf. also Kendall and Stuart (1976, Exercise 3.9). The corresponding formula for vector random variables is obtained as a special case of the result of the next theorem, which contains more general expressions.

THEOREM 4.2. Denote by $\sum_{\underline{v}_m}$ the summation over all ordered positive partitions $v_1 \geq v_2 \geq \dots \geq v_m > 0$ of weight m of v and denote by $\alpha_k = \alpha_k(m) = \#\{v_i = k, i=1, \dots, m\}$. Then

$$\begin{aligned} \mu_{r+v} = & \sum_{m=1}^v \sum_{\underline{v}_m} \sum_{i_1=1}^{\infty} \dots \sum_{i_m=1}^{\infty} \binom{r}{i_m - v}_+ \frac{v!}{\prod_{k=1}^v k!^{\alpha_k} \alpha_k!} \\ & \times S_{-r+v}(\mu_{r+v-i,x} \otimes (\bigotimes_{k=1}^m \kappa_{i_k,x})) \end{aligned}$$

where $i = i_1 + \dots + i_m$ and $\underline{i}_m = (i_1, \dots, i_m)$. □

From Theorem 4.2 we obtain the result for $v=1$:

$$\mu_{r+1,x} = \sum_{i=1}^{r+1} \binom{r}{i-1}_+ S_{-r+1}(\mu_{r+1-i,x} \otimes \kappa_{i,x}).$$

To prove Theorem 4.2 we need the following lemma.

LEMMA 4.1. Denote by $\sum_{m,v}$ the summation over the set $I_{m,v}$ of all non-negative integers $u_{1,s}^m, \dots, u_{m,s}^m$, $s=1, \dots, v$, such that $u_{1,s}^m + \dots + u_{m,s}^m = 1$, $s=1, \dots, v$, $u_{i,j}^m = 0$ for $i > j$ and $|u_{k,v}^m| = u_{k,1}^m + \dots + u_{k,v}^m > 0$. Then it holds that

$$\begin{aligned} & (d^v(\exp \circ \sum_{i=1}^{\infty} \kappa_{i,x}^* \circ \pi_X^{i/i!}))(t_0)(t_1, \dots, t_v) \\ & = ((\exp \circ \sum_{i=1}^{\infty} \kappa_{i,x}^* \circ \pi_X^{i/i!})(t_0)) \\ & \times \sum_{m=1}^v \sum_{i_1=1}^{\infty} \dots \sum_{i_m=1}^{\infty} \sum_{m,v} \frac{1}{\prod_{k=1}^m (i_k - |u_{k,v}^m|)!} \end{aligned}$$

$$\times \left(\prod_{k=1}^m (\kappa_{i_k, x}^* \underline{u}_{k, v}^m, i_k - |\underline{u}_{k, v}^m| \left(\left(\bigotimes_{s=1}^v t_s^{\langle u_{k, s}^m \rangle} \right) \otimes t_0^{\langle i_k - |\underline{u}_{k, v}^m| \rangle} \right) \right)$$

PROOF. Differentiating once, we obtain by aid of Corollary 3.1 that the left hand side mapping of (4.10) has the differential

$$\begin{aligned} & (d(\exp \circ \sum_{i=1}^{\infty} \kappa_{i, x}^* \circ \pi_X^i / i!))(t_0)(t_1) \\ &= ((d \exp) \left(\left(\sum_{i=1}^{\infty} \kappa_{i, x}^* \circ \pi_X^i / i! \right)(t_0) \right) \circ (d \left(\sum_{i=1}^{\infty} \kappa_{i, x}^* \circ \pi_X^i / i! \right)(t_0)))(t_1) \\ &= ((\exp \circ \sum_{i=1}^{\infty} \kappa_{i, x}^* \circ \pi_X^i / i!)(t_0)) \left(\sum_{i=1}^{\infty} i \kappa_{i, x}^* (S_{1, i-1}(t_1 \otimes t_0^{\langle i-1 \rangle})) / i! \right) \end{aligned}$$

which proves that the result is true for $v=1$. Now assume that the result is true for $v=s$. Then we have by aid of Lemma 3.2 that

$$\begin{aligned} & (D(\exp \circ \sum_{i=1}^{\infty} \kappa_{i, x}^* \circ \pi_X^i / i!))_{t_1, \dots, t_s}^{(s)}(t_0)(t_{s+1}) \\ (4.12) \quad &= ((\exp \circ \sum_{i=1}^{\infty} \kappa_{i, x}^* \circ \pi_X^i / i!)(t_0)) \\ & \times \left(\sum_{j=1}^{\infty} \kappa_{j, x}^* (S_{1, j-1}(t_{s+1} \otimes t_0^{\langle j-1 \rangle})) / (j-1)! \right) \\ & \times \sum_{m=1}^s \sum_{i_1=1}^{\infty} \dots \sum_{i_m=1}^{\infty} \sum_{m, s} \frac{1}{\prod_{k=1}^m (i_k - |\underline{u}_{k, s}^m|)!} \\ & \times \prod_{k=1}^m (\kappa_{i_k, x}^* (\underline{u}_{k, s}^m, i_k - |\underline{u}_{k, s}^m| \left(\left(\bigotimes_{j=1}^s t_j^{\langle u_{k, j}^m \rangle} \right) \otimes t_0^{\langle i_k - |\underline{u}_{k, s}^m| \rangle} \right))) \\ & + ((\exp \circ \sum_{i=1}^{\infty} \kappa_{i, x}^* \circ \pi_X^i / i!)(t_0)) \\ & \times \sum_{m=1}^s \sum_{i_1=0}^m \dots \sum_{i_m=0}^m \sum_{m, s} \frac{1}{\prod_{k=1}^m (i_k - |\underline{u}_{k, s}^m| - u_{k, s+1}^m)!} \\ & \times \left(\prod_{k=1}^m (\kappa_{i_k, x}^* (\underline{u}_{k, s+1}^m, i_k - |\underline{u}_{k, s}^m| - u_{k, s+1}^m \left(\left(\bigotimes_{j=1}^{s+1} t_j^{\langle u_{k, j}^m \rangle} \right) \otimes t_0^{\langle i_k - |\underline{u}_{k, s}^m| - u_{k, s+1}^m \rangle} \right))) \right) \end{aligned}$$

where $\sum^m$ denotes the summation over all non-negative integers

$u_{1,s+1}^m, \dots, u_{m,s+1}^m$ such that $u_{1,s+1}^m + \dots + u_{m,s+1}^m = 1$.

Now we have

$$\begin{aligned}
 & \left(\sum_{j=1}^{\infty} \kappa_{j,x}^* (S_{1,j-1} (t_{s+1} \otimes t_0^{\langle j-1 \rangle})) / (j-1)! \right) \\
 & \times \sum_{i_1=1}^{\infty} \dots \sum_{i_{m-1}=1}^{\infty} \sum_{m-1,s} \frac{1}{\prod_{k=1}^{m-1} (i_k - |u_{k,s}^{m-1}|)!} \\
 & \times \prod_{k=1}^{m-1} (\kappa_{i_k,x}^* (S_{u_{k,s}^{m-1}, i_k - |u_{k,s}^{m-1}|}^m ((\otimes_{j=1}^s t_j^{\langle u_{k,j}^{m-1} \rangle}) \otimes t_0^{\langle i_k - |u_{k,s}^{m-1}| \rangle}))) \\
 & + \sum_{i_1=1}^m \dots \sum_{i_m=1}^{\infty} \sum_{m,s} \frac{1}{\prod_{k=1}^m (i_k - |u_{k,s}^m| - u_{k,s+1}^m)!} \\
 & \times \left(\prod_{k=1}^m (\kappa_{i_k,x}^* (S_{u_{k,s+1}^m, i_k - |u_{k,s}^m| - u_{k,s+1}^m}^m ((\otimes_{j=1}^{s+1} t_j^{\langle u_{k,j}^m \rangle}) \otimes t_0^{\langle i_k - |u_{k,s}^m| - u_{k,s+1}^m \rangle}))) \right) \\
 & = \sum_{i_1=1}^{\infty} \dots \sum_{i_m=1}^{\infty} \sum_{m,s+1} \frac{1}{\prod_{k=1}^m (i_k - |u_{k,s+1}^m|)!} \\
 & \times \left(\prod_{k=1}^m (\kappa_{i_k,x}^* (S_{u_{k,s+1}^m, i_k - |u_{k,s+1}^m|}^m ((\otimes_{j=1}^{s+1} t_j^{\langle u_{k,j}^m \rangle}) \otimes t_0^{\langle i_k - |u_{k,s+1}^m| \rangle}))) \right)
 \end{aligned}$$

and hence (4.12) can be written

$$\begin{aligned}
 & ((\exp \circ \sum_{i=1}^{\infty} \kappa_{i,x}^* \circ \pi_X^i / i!)(t_0)) \\
 & \sum_{m=1}^{s+1} \sum_{i_1=1}^{\infty} \dots \sum_{i_m=1}^{\infty} \sum_{m,s+1} \frac{1}{\prod_{k=1}^m (i_k - |u_{k,s+1}^m|)!} \\
 & \times \prod_{k=1}^m (\kappa_{i_k,x}^* (S_{u_{k,s+1}^m, i_k - |u_{k,s+1}^m|}^m ((\otimes_{j=1}^{s+1} t_j^{\langle u_{k,j}^m \rangle}) \otimes t_0^{\langle i_k - |u_{k,s+1}^m| \rangle})))
 \end{aligned}$$

which shows that the result is then also true for $v=s+1$. An induction argument over v now shows that the result is true for all positive

integers. □

PROOF OF THEOREM 4.2. By replacing

$$(\exp \circ \sum_{i=1}^{\infty} \kappa_{i,x}^* \circ \pi_X^i/i!)(t_0)$$

with

$$\sum_{s=0}^{\infty} \langle t_0^{<s>}, \mu_{s,x} \rangle$$

in the expression of Lemma 4.1 we obtain

$$\begin{aligned} & (d^v(\exp \circ \sum_{i=1}^{\infty} \kappa_{i,x}^* \circ \pi_X^i/i!))(t_0)(t_1, \dots, t_v) \\ & \sum_{s=0}^{\infty} \sum_{m=1}^v \sum_{i_1=1}^{\infty} \dots \sum_{i_m=1}^{\infty} \sum_{m,v} \frac{1}{s! \prod_{k=1}^m (i_k - |\underline{u}_{k,v}^m|)!} \\ & (\mu_{s,x}^* \otimes (\bigotimes_{k=1}^m \kappa_{i_k,x}^*)^* \circ (I_X^{<s>} \otimes (\bigotimes_{k=1}^m S_{\underline{u}_{k,v}^m, i_k - |\underline{u}_{k,v}^m|}^m))) \\ & (t_0^{<s>} \otimes (\bigotimes_{k=1}^m ((\bigotimes_{j=1}^v \langle t_j^{<u_{k,j}>} \rangle) \otimes t_0^{<i_k - |\underline{u}_{k,v}^m|>}))) \end{aligned}$$

Now, let $\sum_{v,r}^m$ denote the summation over the set $I_{v,r}^m$ of all non-negative integers $v_{0,j}^m, \dots, v_{m,j}^m$, $j=1, \dots, r$ such that $v_{0,j}^m + \dots + v_{m,j}^m = 1$, $j=1, \dots, r$, $v_{0,1}^m + \dots + v_{0,r}^m = s$ and $v_{k,1}^m + \dots + v_{k,r}^m = i_k - |\underline{u}_{k,r}^m|$, $k=1, \dots, m$.

By Lemma 3.2 we have that (with $i=i_1+\dots+i_m$)

$$\begin{aligned} & (d^r((I_X^{<s>} \otimes (\bigotimes_{k=1}^m (c_{v_k} \otimes I_X^{<i_k - |\underline{u}_{k,v}^m|>}))) \circ \pi_X^{s+i-v})(0_X)(w_1, \dots, w_r) \\ & = \begin{cases} \sum_{v,r}^m s! (\prod_{k=1}^m (i_k - |\underline{u}_{k,v}^m|)!) (\bigotimes_{k=0}^m (v_k \otimes (\bigotimes_{j=1}^m \langle w_j^{<v_{k,j}^m}>))) & s=r+v-i \\ 0_{X^{<s+i-v>}} & s \neq r+v-i \end{cases} \end{aligned}$$

where $v_0 = 1_K$. Writing $u_{k,v+j}^m$ for $v_{k,j}^m$, $k=0, \dots, m$, and t_{v+j} for w_j

$$(d^{v+r}(\exp \circ \sum_{i=1}^{\infty} \kappa_{i,x}^* \circ \pi_X^i/i!))(t_0)(t_1, \dots, t_v, t_{v+1}, \dots, t_{v+r})$$

$$= \sum_{m=1}^v \sum_{i_1=1}^{\infty} \dots \sum_{i_m=1}^{\infty} \sum_{m,v} \sum_{v,r}^m \\ \times (\mu_{r+v-i,x} \otimes (\bigotimes_{k=1}^m \kappa_{i_k,x}))^* (\bigotimes_{k=0}^m (\bigotimes_{j=1}^{v+r} t_j^{<u_{k,j}^m>}))$$

where $u_{0,j}^m = 0$ for $j=1, \dots, v$. Since

$$(d^{v+r}(\exp \circ \sum_{i=1}^{\infty} \kappa_{i,x}^* \circ \pi_X^{i/i!}))(t_0)(t_1, \dots, t_v, t_{v+1}, \dots, t_{v+r})$$

is symmetric in $t_1, \dots, t_{v+r}$, we have

$$(d^{v+r}(\exp \circ \sum_{i=1}^{\infty} \kappa_{i,x}^* \circ \pi_X^{i/i!}))(t_0)(t_1, \dots, t_v, t_{v+1}, \dots, t_{v+r}) \\ = \sum_{m=1}^v \sum_{i_1=1}^{\infty} \dots \sum_{i_m=1}^{\infty} \sum_{m,v} \sum_{v,r}^m \\ \times (\mu_{r+v-i,x} \otimes (\bigotimes_{k=1}^m \kappa_{i_k,x}))^* S_{\frac{1}{-v+r}} (\bigotimes_{k=0}^m (\bigotimes_{j=1}^{v+r} t_j^{<u_{k,j}^m>})) \\ = \sum_{m=1}^v \sum_{i_1=1}^{\infty} \dots \sum_{i_m=1}^{\infty} h (\mu_{r+v-i,x} \otimes (\bigotimes_{k=1}^m \kappa_{i_k,x}))^* S_{\frac{1}{-v+r}} (\bigotimes_{j=1}^{v+r} t_j)$$

where h denotes the number of elements in the summationsets $I_{m,v}$ and $I_{v,r}^m$. The set $I_{v,r}^m$ contains $r!/((r+v-i)! \prod_{k=1}^m (i_k! |u_{k,r}^m|!))$ distinct setups of the integers $v_{0,j}^m, \dots, v_{m,j}^m$, $j=1, \dots, r$ when $|u_{k,r}^m|$ are given and $s=r+v-i$. By taking $t_0 = 0_X$ in Lemma 4.1 we see that

$$\hat{\mu}_{v,x}(t_1, \dots, t_v) = \sum_{m=1}^v \sum_{m,v} (\bigotimes_{k=1}^m \kappa_{i_k})^* S_{\frac{1}{-v}} (\bigotimes_{s=1}^v t_s)$$

where $i_k = |u_{k,v}^m|$. By comparing this expression with the result of Theorem 4.1 (i) we infer that the summation over $I_{m,v}$ can be replaced by a summation over all ordered positive partitions $\frac{v}{-m}$ of weight m of v , and weighted by the factor $v!/(\prod_{k=1}^v k!^{\alpha_k} \alpha_k!)$. This completes the proof of the theorem. $\square$

(c) Differential relations

In the scalar case, moments about zero of certain order can be written in terms of cumulants by differentiating moments of higher order, expressed in terms of cumulants, with respect to a specific cumulant. To be more

precise,

$$\frac{\partial \mu_{r,x}}{\partial \kappa_{s,x}} = \begin{pmatrix} r \\ s \end{pmatrix}_+ \mu_{r-s,x}$$

cf. Kendall and Stuart (1976, eq. (3.35)), or when written as the image of z under the differential at $\kappa_{s,x}$

$$(4.13) \quad (d\mu_{r,x})(\kappa_{s,x})(z) = \begin{pmatrix} r \\ s \end{pmatrix}_+ \mu_{r-s,x} z.$$

The following result generalizes this expression to dimensions larger than 1.

THEOREM 4.3. Consider $\mu_{r,x}$ as the image of $\kappa_{s,x}$ for fixed $\kappa_{i,x}$, $i=1, \dots, s-1, s+1, \dots, r$; then for $s=1, 2, \dots, r$,

$$(d\mu_{r,x})(\kappa_{s,x})(z) = \begin{pmatrix} r \\ s \end{pmatrix}_+ S_{\frac{1}{r}}(\mu_{r-s,x} \otimes z), \quad z \in X^{(s)}.$$

PROOF. From Theorem 3.1 we infer that

$$(d\pi_X^{j_s})(\kappa_{s,x})(z) = j_s S_{j_s-1,1}(\kappa_{s,x}^{\langle j_s-1 \rangle} \otimes z)$$

By aid of Theorem 4.1 and Remark 4.4 we have

$$(4.14) \quad \begin{aligned} (d\mu_{r,x})(\kappa_{s,x})(z) &= \sum C_{\frac{1}{r}-1, \frac{1}{r}} S_{\frac{1}{r}}((\otimes_{i \neq s} \kappa_{i,x}^{\langle j_i \rangle}) \otimes j_s S_{j_s-1,1}(\kappa_s^{\langle j_s-1 \rangle} \otimes z)) \\ &= \sum C_{\frac{1}{r}-1, \frac{1}{r}} S_{\frac{1}{r}} \circ (I_X^{\langle r-s \rangle} \otimes S_{j_s-1,1})(\otimes_{i \neq s} \kappa_{i,x}^{\langle j'_i \rangle} \otimes \kappa_{s,x}^{\langle j'_s \rangle} \otimes z) \end{aligned}$$

where $j'_i = j_i$ if $i \neq s$ and $j'_s = j_s - 1$. Since

$$S_{\frac{1}{r}} \circ (I_X^{\langle r-s \rangle} \otimes S_{j_s-1,1}) = S_{\frac{1}{r}} = S_{\frac{1}{r}} \circ (S_{\frac{1}{r-s}} \otimes I_X^{\langle s \rangle})$$

according to Theorem 2.1 (i), and $\sum_{i=1}^r i j'_i = \sum_{i=1}^r i j_i - s = r-s$, it follows that (4.14) can be written

$$\begin{aligned} \frac{r!}{(r-s)!s!} \sum_{i=1}^r \frac{(r-s)!}{\prod (j'_i! i!^{j'_i})} S_{\frac{1}{r}}(S_{\frac{1}{r-s}}(\otimes_{i=1}^r \kappa_{i,x}^{\langle j'_i \rangle}) \otimes z) \\ = \begin{pmatrix} r \\ s \end{pmatrix}_+ S_{\frac{1}{r}}(\mu_{r-s,x} \otimes z) \end{aligned}$$

which proves the theorem. $\square$

When $\dim X = 1$ the result of Theorem 4.3 reduces to (4.13) as it should.

4.4. Some properties

In order to show some properties of the moments and cumulants we need the following lemma.

LEMMA 4.2. Let $\psi_x \in D^r(T;K)$ denote either the m.g.f. or the c.g.f. of the random variable $x \in X$, and let $\gamma_{r,x}$ denote the corresponding characteristic of order r of x , that is the r 'th moment about zero or the r 'th cumulant. Let $A \in L(X;V)$; then

$$\hat{\gamma}_{r,Ax} = (d^r(\psi_x \circ A^*)) (0_V)$$

PROOF. Since $\langle s, Ax \rangle_V = \langle A^* s, x \rangle_X$, for $s \in V$, it follows that $\psi_{Ax}(s) = \psi_x(A^* s) = (\psi_x \circ A^*)(s)$, and hence the result follows from (4.3) and (4.4). $\square$

The following theorem gives concise forms of some properties of invariance kind for moments and cumulants.

THEOREM 4.4. Let $a \in V$ and $A \in L(X;V)$ then for the random variable $x \in X$,

$$(i) \quad \mu_{r,Ax} = A^{\langle r \rangle} \mu_{r,x} \quad r=1,2,\dots$$

$$(ii) \quad \kappa_{1,Ax+a} = A \kappa_{1,x} + a$$

$$(iii) \quad \kappa_{r,Ax+a} = A^{\langle r \rangle} \kappa_{r,x} \quad r=2,3,\dots$$

PROOF. It is easily verified that $C_{Ax+a} = k a^* + C_{Ax}$, where $k = \langle 1_K, 1_K \rangle_K$, and hence it follows by Lemma 3.1 (ii) that

$$\hat{\kappa}_{1,Ax+a} = (d(k a^* + C_{Ax})) (0_V) = k a^* + (dC_{Ax}) (0_V) = k a^* + \hat{\kappa}_{1,Ax}$$

while

$$\hat{\kappa}_{r,Ax+a} = (d^r(k a^* + C_{Ax})) (0_V) = (d^r C_{Ax}) (0_V) = \hat{\kappa}_{r,Ax}$$

for $r > 1$. It thus remains to show that $\gamma_{r,Ax}^* = \gamma_{r,x}^* \circ A^{\langle r \rangle*}$ for γ^*

being μ^* or κ^* . By combining Lemma 4.2 and (3.2) we obtain

$$\begin{aligned}\gamma_{r,Ax}^* &= L_{\odot}^{-1}(\hat{\gamma}_{r,Ax}) = L_{\odot}^{-1}((d^r(\psi_x \circ A^*)) (0_Y)) = L_{\odot}^{-1}((d^r(L_A^*(\psi_x))) (0_Y)) \\ &= L_{\odot}^{-1}(d^r \psi_x (A^* 0_Y)) \circ A^{*\langle r \rangle} = \gamma_{r,x}^* \circ A^{*\langle r \rangle} = \gamma_{r,x}^* \circ A^{\langle r \rangle*}. \quad \square\end{aligned}$$

By a polynomial in x of degree k we mean the expression

$$(4.15) \quad y = A_0 + A_1 x + A_2 x^{\langle 2 \rangle} + \dots + A_k x^{\langle k \rangle}$$

where $A_i \in L(X^{\langle i \rangle}; Y)$, $i=0,1,\dots,k$. The following theorem expresses the moments and cumulants of a polynomial in x in terms of the moments and cumulants of x . For convenience we use the following notation: Let $S_{\underline{r}_k}^Y$ be the symmetrizer on $Y^{\langle r \rangle}$ connected with the partition $\underline{r}_k$ of r and let

$$B_{\underline{m},s}^r = \sum_{\underline{m}_k}^s \binom{r}{\underline{m}_k} + S_{r-\underline{m},\underline{m}_k}^Y \circ (A_0^{\langle r-\underline{m} \rangle} \otimes (\bigotimes_{i=1}^k A_i^{\langle m_i \rangle}))$$

and

$$C_s^r = \sum_{m=1+\lfloor \frac{s-1}{k} \rfloor}^r B_{m,s}^r.$$

where $[x]$ is the integer part of x and $\sum_{\underline{m}_k}^s$ is defined in Theorem 4.1.

THEOREM 4.5. Let y be given by (4.15). Then the following results hold.

$$(i) \quad \mu_{r,y} = A_0 + \sum_{s=1}^{kr} \sum_{m=1+\lfloor \frac{s-1}{k} \rfloor}^r B_{m,s}^r \mu_{s,x} = A_0 + \sum_{s=1}^{kr} C_s^r \mu_{s,x}$$

for $r = 1, 2, \dots$.

(ii) Especially for $A_0 = c_{0_Y} \in M(K; Y)$ we have

$$\mu_{r,y} = \sum_{s=1}^{kr} B_{r,s}^r \mu_{s,x}$$

for $r = 1, 2, \dots$.

$$(iii) \quad \kappa_{1,y} = \mu_{1,y} = A_0 + A_1 \mu_{1,x} + A_2 \mu_{2,x} + \dots + A_k \mu_{k,x}$$

$$(iv) \quad \kappa_{r,y} = \sum_{j=1}^r \sum_{\underline{j}_r}^r (-1)^{j-1} (j-1)! C_{\underline{j}_r}^r S_{\underline{j}_r}^Y \left(\bigotimes_{i=1}^r \left(\sum_{s=1}^{ki} B_{i,s}^i \mu_{s,x} \right)^{\langle j_i \rangle} \right)$$

$$= \sum_{j=1}^r \sum_{\underline{j}_r}^r (-1)^{j-1} (j-1)! \cdot \\ \times C_{\underline{j}_r} S_{\underline{j}_r}^y \left(\bigotimes_{i=1}^r \left(\sum_{s=1}^{ki} B_{i,s}^i \sum_{v=1}^s \sum_{\underline{v}_s}^s C_{\underline{v}_s} S_{\underline{v}_s} \left(\bigotimes_{p=1}^s \kappa_{p,x}^{<v>} \right)^{<j_i>} \right) \right)$$

for $r=2,3,\dots$.

PROOF. (i): We have by (4.7) and the multinomial expansion (2.6)

$$\mu_{r,y} = E(y^{<r>}) = E\left(\sum_{i=0}^k A_i x^{<i>} \right)^{<r>} \\ = E\left(\sum_{m=0}^r \sum_{\underline{m}_k} \binom{r}{\underline{m}_k} (S_{r-m,\underline{m}_k}^y \circ (A_0^{<r-m>} \otimes \left(\bigotimes_{i=1}^k A_i^{<m_i>}\right)) \left(\bigotimes_{i=1}^k x^{<im_i>}\right)\right) \\ = \sum_{m=0}^r \sum_{s=0}^{km} \sum_{\underline{m}_k}^s \binom{r}{\underline{m}_k} (S_{r-m,\underline{m}_k}^y \circ (A_0^{<r-m>} \otimes \left(\bigotimes_{i=1}^k A_i^{<m_i>}\right))) \mu_{s,x}$$

By interchanging the order of summation over m and s we obtain

$$\mu_{r,y} = A_0 + \sum_{s=1}^{kr} \sum_{m=1+}^r \left[\frac{s-1}{k} \right] B_{m,s}^r \mu_{s,x} = A_0 + \sum_{s=1}^{kr} C_s^r \mu_{s,x}$$

(ii): Since $B_{m,s}^r = 0_{<r>}$ for $m < r$ when $A_0 = c_{0,y}$, the result follows immediately from (i).

(iii): We have

$$\kappa_{1,y} = E(A_0 + A_1 x + \dots + A_k x^{<k>}) = A_0 + A_1 \mu_{1,x} + \dots + A_k \mu_{k,x}.$$

(iv): We have from Theorem 4.4 (iii) that $\kappa_{r,y} = \kappa_{r,y-A_0}$ for $r=2,3,\dots$

Let $y' = y - A_0$; then by (ii) and Theorem 4.1 (iii)

$$\kappa_{r,y'} = \sum_{j=1}^r \sum_{\underline{j}_r}^r (-1)^{j-1} (j-1)! C_{\underline{j}_r} S_{\underline{j}_r}^y \left(\bigotimes_{i=1}^r \mu_{i,y'}^{<j_i>} \right) = \\ = \sum_{j=1}^r \sum_{\underline{j}_r}^r (-1)^{j-1} (j-1)! C_{\underline{j}_r} S_{\underline{j}_r}^y \left(\bigotimes_{i=1}^r \left(\sum_{s=1}^{ki} B_{i,s}^i \mu_{s,x} \right)^{<j_i>} \right) \\ = \sum_{j=1}^r \sum_{\underline{j}_r}^r (-1)^{j-1} (j-1)! \\ \times C_{\underline{j}_r} S_{\underline{j}_r}^y \left(\bigotimes_{i=1}^r \left(\sum_{s=1}^{ki} B_{i,s}^i \sum_{v=1}^s \sum_{\underline{v}_s}^s C_{\underline{v}_s} S_{\underline{v}_s} \left(\bigotimes_{p=1}^s \kappa_{p,x}^{<v>} \right)^{<j_i>} \right) \right)$$

□

Generalized cumulants for polynomial transformations of the elements of vector valued random variables, expressed in terms of generalized

cumulants of the original variables has been given by McCullagh (1984) using terminology from graph theory. This is also the essence of the main result of Leonov and Shiryaev (1959) where the product-cumulants of the transformed variables are stated in terms of the product-cumulants of the original variables.

Now we shall assume that $X = V \oplus V^\perp$ and that $M_x \in D_V^r(\mathcal{D}; K)$. We define the random variable $x_V: \Omega \rightarrow V$, in terms of the random variable x , by $x_V(\omega) = P_V x(\omega)$ for $\omega \in \Omega$. The m.g.f. of x_V is given by

$$M_{x_V}(t_V) = M_x(t_V)$$

for $t_V \in V$, that is $M_{x_V} = M_x|_V$. Hence

$$\hat{\mu}_{r, x_V} = (d_V^r M_x)(0_X).$$

Similarly we have

$$\hat{\kappa}_{r, x_V} = (d_V^r C_x)(0_X).$$

The next theorem shows the effect of calculating partial differentials of differentiable generating functions.

THEOREM 4.6. Let $M_x \in D^r(\mathcal{D}; K)$ where $0_X \in \mathcal{D}$, and $\mathcal{D}$ is an open subset of T . Then

$$(i) \quad \mu_{r, x_V} = P_V^{\langle r \rangle} \mu_{r, x}$$

$$(ii) \quad \kappa_{r, x_V} = P_V^{\langle r \rangle} \kappa_{r, x}.$$

PROOF. The result immediately follows from Theorem 4.4 by letting $V = V$ and $A = P_V$. □

We shall also consider so called extended generating functions. These are defined in consistency with arbitrary extended mappings discussed in section 3. Thus with $T' = T \oplus W$ we define the extended m.g.f. and the extended c.g.f. by

$$(4.16) \quad M'_x(t') = M_x(P_X t')$$

and

$$(4.17) \quad C'_x(t') = C_x(P_X t')$$

for $t' \in T'$. The notion of "extension" of generating function is just a matter of the definition of the image space of x . If $X = X_1 \oplus X_2$ and $x \in X_1$ almost surely with respect to ν_x we may regard x as a mapping from Ω to X_1 . The generating functions are then defined on a subset T_1 of X_1 , while the functions (4.1) and (4.2) are the corresponding extensions to $T = T_1 \oplus X_2$. The generating functions given by (4.16) and (4.17) are thus the generating functions for the random variable

$$x': \Omega \rightarrow X' = X \oplus W$$

defined by $x'(\omega) = x(\omega)$ for $\omega \in \Omega$ which has singular measure ν_x , on X' such that $x' \in X$ almost surely with respect to ν_x .

The next theorem shows the effect of calculating the differential of the extended generating functions.

THEOREM 4.7. Let x be a random variable on $X \subset X' = X \oplus W$ with m.g.f. $M_x \in D^r(\mathcal{D}; K)$ and c.g.f. $C_x \in D^r(\mathcal{D}; K)$ defined on $T \subset X$ where $0_X \in \mathcal{D} \subset T$. Then

$$(i) \quad (d^r M'_x)(0_{X'}) = \hat{\mu}_{r, P_X^* x},$$

$$(ii) \quad (d^r C'_x)(0_{X'}) = \hat{\kappa}_{r, P_X^* x}.$$

PROOF. From (3.3) we see that

$$\begin{aligned} L_{\odot}^{-1}((d^r \psi'_x)(0_{X'})) &= L_{\odot}^{-1}((d^r \psi_x)(0_X)) \circ P_X^{<r>} = \gamma_{r, x}^* \circ P_X^{<r>} \\ &= (P_X^{<r>*} \circ \gamma_{r, x})^* = \gamma_{r, P_X^* x}^* \end{aligned}$$

where the last equality follows from Theorem 4.4. □

If W in Theorem 4.7 is taken to be the orthogonal complement of X in X' , so that P_X is the orthogonal projection on X and hence hermitian

($P_X^* = P_X$) it follows that $P_X^* x = x$ (since $x \in X$) and the characteristics of x are then also obtained by differentiating the extended mappings.

5. COORDINATE SPACE DIFFERENTIALS: GRADIENTS

5.1. Introduction

Although coordinate free differentials, as those in section 3, are convenient for analysis, coordinate representations of differentials (basis dependent) are convenient for computation (for example computer aided) and also for analysis in cases when the mapping itself has a representation in coordinates as for example as a matrix, or when the argument of the mapping is represented in coordinates. Examples of the latter case is the likelihood with coordinate representations of the parameters such as the mean vector and dispersion matrix in the normal distribution, and the moment generating function of matrix variates depending on matrix arguments.

In section 5.2 we give rules and representations of the differential of a mapping with respect to an arbitrary given basis in a general coordinate space. In section 5.3 we consider differential realizations in some specific coordinate spaces such as the spaces of column m -tuples ("column vectors") and the space of m by n matrices. In section 5.4 we consider the problem of giving representations of differentials in the space of m -square matrices, of mappings with domain in the space of symmetric matrices but where the argument is represented in the basis of the space of m -square matrices. In section 5.5 we generalize this to mappings with domain on a subspace of a space in which the argument has a representation.

5.2. General coordinate space differentials

Let X and Y be separable Hilbert spaces over the same field K and with orthogonal bases $X_b = \{x_i^b \in X, i \in I\}$ and $Y_b = \{y_j^b \in Y, j \in J\}$, respectively. For $x_0 = \sum_{i \in I} x_{i0}^b x_i$ where $x_{i0}^b \in K, i \in I$, we define a differential operator $\partial/\partial x_0$ by

$$\partial/\partial x_0 = \sum_{i \in I} (\partial/\partial x_{i0}^b) x_i^{b*}.$$

Here the set $X_b^* = \{x_i^{b*} \in X^*, i \in I\}$ is the basis of X^* adjoint to X_b with respect to the inner product of X , that is,

$$x_i^{b*}(x_j^b) = \delta_{ij}.$$

Let $y_0 = \sum_{j \in J} y_{j0} y_j^b \in Y$, $y_{j0} = f_j(x_0) \in K$, $j \in J$, be the image $f(x_0)$ of x_0 under the mapping $f \in D^r(X; Y)$. We denote by

$$(5.1) \quad \partial/\partial x_0 \otimes y_0 \equiv \partial/\partial x_0 \otimes f(x_0)$$

the linear mapping

$$\sum_{i \in I} \sum_{j \in J} (\partial y_{j0} / \partial x_{i0}) x_i^{b*} \otimes y_j^b \equiv \sum_{i \in I} \sum_{j \in J} (\partial f_j(x_0) / \partial x_{i0}) x_i^{b*} \otimes y_j^b$$

in $L(X; Y) = X^* \otimes Y$ which we call the gradient of f at x_0 .

More generally we denote by

$$(5.2) \quad (\partial/\partial x_0)^{<r>} \otimes f(x_0) \equiv \left(\bigotimes_{i=1}^r (\partial/\partial x_0) \right) \otimes f(x_0)$$

the mapping

$$\sum_{i_1 \in I} \dots \sum_{i_r \in I} \sum_{j \in J} (\partial^r f_j(x_0) / \partial x_{i_1 0} \dots \partial x_{i_r 0}) x_{i_1}^{b*} \otimes \dots \otimes x_{i_r}^{b*} \otimes y_j^b$$

in $L(X^{(r)}; Y)$ and call this mapping the r 'th order gradient of f at x_0 .

This mapping is also the image, under the natural isomorphism between $L(X; L(X^{(r-1)}; Y))$ and $L(X^{(r)}; Y)$, of

$$(5.3) \quad \partial/\partial x_0 \otimes ((\partial/\partial x_0)^{<r-1>} \otimes f(x_0)).$$

We shall not distinguish between the mapping (5.2) and the corresponding recursively defined mapping (5.3).

The following theorem shows that the r 'th order gradient is a coordinate space representation of the r 'th differential.

THEOREM 5.1. Let $f \in D^r(X; Y)$ and X_b an orthonormal basis of X . Then the r 'th order gradient at $x_0 \in X$, $(\partial/\partial x_0)^{<r>} \otimes f(x_0)$ is a coordinate space representation of the r 'th differential $L_{\odot}^{-1}((d^r f)(x_0))$

PROOF. Let $x_0 = \sum_{i \in I} x_{i0} x_i^b$ and $x_1 = \sum_{i \in I} x_{i1} x_i^b$, then

$$\begin{aligned} (\partial/\partial x_0 \otimes f(x_0))(x_1) &= \sum_{k \in I} \sum_{i \in I} \sum_{j \in J} x_{ki} (\partial \delta_j(x_0)/\partial x_{i0}) x_i^{b*}(x_k^b) y_j^b \\ &= \sum_{i \in I} \sum_{j \in J} x_{i1} (\partial \delta_j(x_0)/\partial x_{i0}) y_j^b \end{aligned}$$

and thus

$$\begin{aligned} &\| f(x_0 + x_1) - f(x_0) - (\partial/\partial x_0 \otimes f(x_0))(x_1) \|_Y \\ &= \left\| \sum_{j \in J} (\delta_j(x_0 + x_1) - \delta_j(x_0) - \sum_{i \in I} x_{i1} (\partial \delta_j(x_0)/\partial x_{i0}) y_j^b) \right\|_Y \\ &\leq \sum_{j \in J} \left\| \delta_j(x_0 + x_1) - \delta_j(x_0) - \sum_{i \in I} x_{i1} (\partial \delta_j(x_0)/\partial x_{i0}) \right\|_K \| y_j^b \|_Y. \end{aligned}$$

Since, for each $j \in J$, $\left\| \delta_j(x_0 + x_1) - \delta_j(x_0) - \sum_{i \in I} x_{i1} (\partial \delta_j(x_0)/\partial x_{i0}) \right\|_K = o(\| x_1 \|_X)$ it follows that

$$\| f(x_0 + x_1) - f(x_0) - (\partial/\partial x_0 \otimes f(x_0))(x_1) \|_Y = o(\| x_1 \|_X)$$

and hence by (3.1)

$$\partial/\partial x_0 \otimes f(x_0) = \left[(Df)(x_0) \right]_{X_b}^{Y_b}.$$

This proves the statement for $r = 1$, since $(Df)(x_0) = (df)(x_0)$ and $L_{\odot}^{-1}((df)(x_0)) = (df)(x_0)$.

From this result we obtain

$$\partial/\partial x_0 \otimes (D^{r-1}f)(x_0) = (D(D^{r-1}f))(x_0) = (D^r f)(x_0)$$

which shows that the recursively defined r 'th order gradient at x_0 can be written

$$(D^r f)(x_0) = \partial/\partial x_0 \otimes (\partial/\partial x_0 \otimes \dots (\partial/\partial x_0 \otimes f(x_0)) \dots).$$

Hence by the isomorphism between $(D^r f)(x_0)$ and $(d^r f)(x_0)$ we have

$$(d^r f)(x_0) = ((\partial/\partial x_0)^{\langle r \rangle} \otimes f(x_0)) \circ \otimes^r.$$

Since

$$((\partial/\partial x_0)^{\langle r \rangle} \otimes f(x_0)) \left(\bigotimes_{i=1}^r x_i \right) = ((\partial/\partial x_0)^{\langle r \rangle} \otimes f(x_0)) \circ P_r^{\otimes}(\sigma) \left(\bigotimes_{i=1}^r x_i \right)$$

for any $\sigma \in S_{1-r}$ and $x_i \in X$, $i \in I$, it follows that

$$((\partial/\partial x_0)^{\langle r \rangle} \otimes f(x_0)) \circ S_{\underline{1}-r} = (\partial/\partial x_0)^{\langle r \rangle} \otimes f(x_0)$$

and hence

$$\begin{aligned} (d^r f)(x_0) &= ((\partial/\partial x_0)^{\langle r \rangle} \otimes f(x_0)) \circ S_{\underline{1}-r} \circ \otimes^r \\ &= L_{\odot}((\partial/\partial x_0)^{\langle r \rangle} \otimes f(x_0)). \end{aligned}$$

This proves the theorem. □

Remark 5.1. The notations (5.1) and (5.2) are not the best choices since if $g(x_0) = f(x_0)$, then from a notational point of view $(\partial/\partial x_0)^{\langle r \rangle} \otimes g(x_0)$ is equal to $(\partial/\partial x_0)^{\langle r \rangle} \otimes f(x_0)$ but the mappings which they represent are not necessarily identical. The reasons for the chosen notation are two. The first reason is that the notation gives an indication how to calculate the mapping. However this could also been achieved by writing $(1/\partial x_0)^{\langle r \rangle} \otimes (\partial^r f)(x_0)$ instead of (5.2) and this would have eliminated the notational confusion. The second reason, which has been the conclusive one, is that the notation (5.1) as well as the use of the differential operator $\partial/\partial x_0$ have become commonly used; cf. Example 5.3 below and the references given there. □

5.3. Matrix derivatives

The notion matrix derivatives has become a comprehensive term for an arrangement of partial derivatives of the elements of a matrix Y with respect to the elements of a matrix X . The cases when Y is scalar or vector valued and X is a vector are of course special cases of matrix derivatives, but these cases are often referred to as "derivatives with respect to vectors", cf. Graybill (1969, p. 261) and Nel (1980, sec. 3), while the term matrix derivatives is used when Y is scalar, vector or matrix valued and X is a matrix or vice versa. The real problem how to arrange the partial derivatives begins when Y is a vector or matrix and

X is a matrix or vice versa.

EXAMPLE 5.1. Let $f \in M(R^m; R)$, where R^m is the space of column m -tuples over R . In this case it is common to define (cf. e.g. Graybill (1969, p. 261) and Rao (1973, p. 71))

$$\partial f / \partial x_0 = (\partial f / \partial x_{10}, \dots, \partial f / \partial x_{m0})^t$$

and

$$(5.4) \quad \partial^2 f / \partial x_0^2 = (\partial^2 f / \partial x_{i0} \partial x_{j0}), \quad i, j = 1, \dots, m$$

for $x_0^t = (x_{10}, \dots, x_{m0})$, where x_0^t denotes the transpose of x_0 . We have here for a moment used the notation used in the references which write ∂f instead of $\partial f(x_0)$.

This way of defining the first order derivative makes it contravariant and not covariant with the vector as it should be. It would be more natural to define the first order derivative as a row vector

$$(5.5) \quad (\partial f(x_0) / \partial x_{10}, \dots, \partial f(x_0) / \partial x_{m0})$$

which is nothing but (5.1) with X_b taken to be the canonical basis $\{E_i, i=1, \dots, m\}$ of R^m , that is, E_i is the column m -tuple with 1 in position i as the only non-zero element. □

EXAMPLE 5.2. When $f \in M(R^m; R^n)$, $x_0 = (x_{10}, \dots, x_{m0})^t \in R^m$ and $y_0 = f(x_0) = (y_{10}, \dots, y_{n0})^t \in R^n$, it is common to let the matrix $(\partial y_{i0} / \partial x_{j0})$, $i=1, \dots, n, j=1, \dots, m$, be a representation of the derivative of f with respect to x_0 . This matrix can formally be written

$$(5.6) \quad (\partial / \partial x_0) \otimes y_0 \equiv (\partial / \partial x_0) \otimes f(x_0)$$

where $\partial / \partial x_0 = (\partial / \partial x_{10}, \dots, \partial / \partial x_{m0})$ and where $\otimes$ denotes the right Kronecker product defined by $A \otimes B = (A b_{i,j})$ for $B = (b_{i,j})$. In this notation we can write (5.4) as

$$\partial^2 f / \partial x_0^2 = (\partial / \partial x_0) \otimes (\partial f / \partial x_0)^t.$$

□

EXAMPLE 5.3. Let $f \in M(R^{m,n}; R^{r,s})$ where $R^{m,n}$ is the space of real m by n matrices. In the case when a matrix $Y = f(X) \in R^{r,s}$ is to be differentiated with respect to the elements of a matrix $X \in R^{m,n}$, has considered by many authors. There are mainly two commonly used definitions of, so called matrix derivatives. One of these, due to MacRae (1974), (cf. also Rogers (1980) and Nel (1980)) is obtained from (5.1) by letting

$$(5.7) \quad \partial/\partial X \otimes Y$$

be realized by the Kronecker product of $\partial/\partial X = (\partial/\partial x_{i,j})$, for $X = (x_{i,j})$, $i=1, \dots, m$, $j=1, \dots, n$, and Y . This is equivalent to letting X_b in (5.1) be the canonical basis $E_{i,j}$, $i=1, \dots, m$, $j=1, \dots, n$ of $R^{m,n}$, that is, $E_{i,j}$ is the m by n matrix with 1 in position (i,j) as the only non-zero element. The dimensions of $E_{i,j}$ shall always be understood from the ranges of i and j .

Another definition, which is considered by Neudecker (1968) among others, is obtained by letting (5.1) be realized as

$$(5.8) \quad \text{vec}_{m,n}(\partial/\partial X)(\text{vec}_{r,s}(Y))^t$$

The $\text{vec}_{m,n}$ operator from $K^{m,n}$ to K^{mn} is here taken to be the isomorphic mapping which stacks the columns of $X \in K^{m,n}$ on top of another into a longer column, that is, $\text{vec}_{m,n}(X) = (x_{\cdot 1}^t, \dots, x_{\cdot n}^t)^t \in K^{mn}$ where $x_{\cdot j} = (x_{1,j}, \dots, x_{m,j})^t$ denotes the j 'th column of X . When m and n are to be understood from the order of the argument X , we write $\text{vec}(X)$ for $\text{vec}_{m,n}(X)$.

McDonald and Swaminathan (1973) and Swaminathan (1976) consider a realization which in our notation can be written

$$(5.9) \quad \text{vec}_{n,m}(\partial/\partial X^t)(\text{vec}_{s,r}(Y^t))^t;$$

Tracy and Dwyer (1969) representations similar to (5.8) and (5.9), but in another notation. □

We shall use a representation of (5.1) (Nel (1980, Def. 5.2.1)) which

can be written as

$$(5.10) \quad \text{vec}_{r,s}(Y)(\text{vec}_{m,n}(\partial/\partial X))^t.$$

The reason for writing the differential operator to the right of the operand is that we want to have the differential to be covariant with the argument on which it acts. Also the chain-rule will then have a very simple formula in terms of ordinary matrix multiplication and with inner differentials written to the right similar to what is common in the scalar case.

The representation (5.10) can also be written

$$(\text{vec}_{m,n}(\partial/\partial X))^t \otimes \text{vec}_{r,s}(Y)$$

where, here and in the sequel $\otimes$ denotes the right Kronecker product defined as in Example 5.2. This realization is conformable with (5.6) for vector derivatives.

We shall use the notation $D_X \otimes Y$ for (5.7) and $d_X \square Y$ for (5.10). The relations between these are

$$(5.11) \quad d_X \square Y = D_{\text{vec}_{m,n}(X)} \otimes \text{vec}_{r,s}(Y) = (\text{vec}_{m,n}(D_X))^t \otimes \text{vec}_{r,s}(Y).$$

As seen from these examples, in our functional analytical vocabulary, matrix derivatives are differentials of a mapping at a certain point and are themselves linear mappings. They are not derivatives in a certain direction in the functional analytical sense.

Since the matrix derivatives are linear mappings, it is natural in Examples 5.1 and 5.2 to have them realized as matrices, as in (5.5) and (5.6) since linear mappings from K^m to K^n are realized as matrices. In Example 5.3 however there is no argument for representing the matrix derivative as a matrix as done both in (5.7) and (5.8). This because linear mappings from $K^{m,n}$ to $K^{r,s}$ cannot be represented as matrices if the ordinary multiplication rule is to be maintained. However, for a rs by mn matrix A ,

$$\text{vec}_{r,s}^{-1} \circ A \circ \text{vec}_{m,n}$$

is a linear mapping from $K^{m,n}$ to $K^{r,s}$ isomorphic to A . Hence the arrangement (5.8) (5.9) and (5.10) can all be motivated as matrices isomorphic to the linear mapping from $K^{m,n}$ to $K^{r,s}$ which is the true differential at the specific point X . Since there exists a bijective mapping between $d_X \square Y$ and $D_X \otimes Y$, the same is true for $D_X \otimes Y$. The image of a point X_1 under the differential at X , that is the derivative of f at X in the direction X_1 , can be calculated as

$$\text{vec}_{r,s}^{-1}((d_X \square f(X))\text{vec}_{m,n}(X_1))$$

by aid of (5.10); here the juxtaposition in the argument of $\text{vec}_{r,s}^{-1}$ is ordinary matrix multiplication. This resulting matrix in $K^{r,s}$ can also be obtained by aid of (5.7) although the expression is somewhat more complicated.

Thus it is better to use the term "matrix derivatives" as a comprehensive term for matrix representations of "derivatives" than as "derivatives" of matrices.

Higher order "derivatives" defined in a recursive manner as "derivatives" of "derivatives", cf. e.g. Rogers (1980, Def. 10.3), are in our notion higher order differentials at a certain point, and are thus multilinear mappings. By use of (5.4) it is possible to write the second order derivative of f at x_0 in directions $x_1 \in K^m$ and $x_2 \in K^m$, that is the image under the second order differential at x_0 , as

$$(5.12) \quad x_1^t (\partial^2 f / \partial x_0^2) x_2 \in K$$

where the juxtaposition is ordinary matrix multiplication. However, notational problems arise when we look at the third order derivative of f at x in directions $x_1 \in K^m$, $x_2 \in K^m$ and $x_3 \in K^m$. It is clear that the third order differential of f at x_0 can be written as the box array

$$\partial^3 f(x_0) / \partial x_0^3 \equiv (\partial^3 f(x_0) / \partial x_{i0} \partial x_{j0} \partial x_{k0}), \quad i, j, k = 1, \dots, m$$

and thus, with $x_i^t = (x_{1i}, \dots, x_{mi})$, $i=1,2,3$, we can write the third order derivative of f at x_0 in directions x_1, x_2 and x_3 as

$$\sum_{ijk} (\partial^3 f(x_0) / \partial x_{i0} \partial x_{j0} \partial x_{k0}) x_{i1} x_{j2} x_{k3}.$$

However this notation requires the coordinates of x_1, x_2, x_3 and $\partial^3 f(x_0) / \partial x_0^3$, and is thus not fully as convenient as the notation (5.12) for the second order derivative.

This notational problem arises at an earlier stage in Examples 5.2 and 5.3 and are all examples of the problem of how to represent multilinear mappings from coordinate spaces to coordinate spaces.

As seen from the examples, the realization of the differential at a certain point, depends on the particular choice of basis as well as the particular choice of realization of the tensor product. A variety of different (but isomorphic) representations can thus be defined. As long as one is solely interested in the differential at a specific point as a mapping, the particular representation, that is, the arrangement of the partial derivatives, is immaterial from an analytical point of view.

However when the image of some specific arguments under this mapping, that is the derivative in certain directions at the specific point, is to be calculated, it is essential to know how the specific arrangement acts on its arguments. In this case and also in the case when differentials of complicated expressions are to be calculated by aid of certain rules as for example the chain-rule, then some realizations may have advantages before others from an algebraic point of view.

The representations (5.6) and (5.10) are special cases of a class of representations given by the array

$$(5.13) \quad d_X \square Y = \begin{pmatrix} \frac{\partial y_{i(1)}}{\partial x_{j(1)}} & \dots & \frac{\partial y_{i(1)}}{\partial x_{j(m)}} \\ \vdots & & \vdots \\ \frac{\partial y_{i(n)}}{\partial x_{j(1)}} & \dots & \frac{\partial y_{i(n)}}{\partial x_{j(m)}} \end{pmatrix}$$

where $m=\dim X$ and $n=\dim V$ and where $X = (x_{\underline{i}_u})$ and $Y = (y_{\underline{j}_v})$ are hypermatrix representations in u and v dimensions respectively. Here the mappings $i: \{1,2,\dots,n\} \rightarrow J$ and $j: \{1,2,\dots,m\} \rightarrow I$ are one-to-one. For mappings between column vector spaces ($u=v=1$) as in Example 5.2, $I = \{1,2,\dots,m\}$ and $J = \{1,2,\dots,n\}$ and both i and j can be taken as identity mappings which result in the representation (5.6). For mappings between matrix spaces ($u=v=2$) as in Example 5.3, $I = \{(\alpha,\beta), \alpha=1,\dots,m, \beta=1,\dots,n\}$ and $J = \{(\alpha,\beta), \alpha=1,\dots,r, \beta=1,\dots,s\}$ and we may take i and j to be the lexicographical orderings of the sets which corresponds to letting

$$i(k) = (1+((k-1) \bmod m), 1 + \lfloor k/m \rfloor).$$

where $\lfloor x \rfloor$ denotes the integer part of x , and similarly for j .

For hypermatrix spaces we introduce the following notations.

Denote by $I(\underline{n}_u)$ the totality of $n_1 \cdots n_u$ u -tuples $(i_1, \dots, i_u) = \underline{i}_u$ of positive integers which satisfy $1 \leq i_k \leq n_k, k=1, \dots, u$ where $\underline{n}_u = (n_1, \dots, n_u)$. We define a mapping

$$p: I(\underline{n}_u) \rightarrow I(n_1 \cdots n_u)$$

by

$$p(\underline{i}_u) = 1 + \langle \underline{i}_u - \underline{1}_u, \underline{\Pi}_u(\underline{n}_u) \rangle \equiv 1 + \sum_{k=1}^u (i_k - 1) \Pi_{k-1}(\underline{n}_u)$$

where $\underline{1}_u$ is a u -tuple of ones, $\underline{\Pi}_u(\underline{n}_u) = (\Pi_0(\underline{n}_u), \dots, \Pi_{u-1}(\underline{n}_u))$, $\Pi_k(\underline{n}_u) = n_0 n_1 \cdots n_k, k=0, 1, \dots, u-1$ and $n_0=1$. Of course, p is merely a family of mappings parameterized by r and $n_1, \dots, n_u$. These parameters should be understood from the ranges of $i_1, \dots, i_u$, but when we want to emphasize these parameters, we write $p(\underline{i}_u; \underline{n}_u)$ for the image of $\underline{i}_u$ under the mapping p with parameters $\underline{n}_u$.

The inverse $p^{-1}(\cdot; \underline{m}_u)$ is given by

$$p^{-1}(k; \underline{m}_u) = (i_1, \dots, i_u)$$

where

$$i_k = 1 + \lfloor ((k-1) \bmod (m_0 \cdots m_k)) / (m_0 \cdots m_{k-1}) \rfloor.$$

By aid of the mapping p we can write the vec mapping, given below

expression (5.8), as

$$\text{vec}_{m,n}(X) = \sum_{i=1}^m \sum_{j=1}^n x_{i,j} E_{p(i,j;m,n)} \in \mathbb{R}^{mn}$$

for

$$X = \sum_{i=1}^m \sum_{j=1}^n x_{i,j} E_{i,j} \in \mathbb{R}^{m,n}.$$

This vec mapping is a special case of a class of vec mappings, from hypermatrix spaces to columnvector spaces, defined by aid of the mapping p , cf. Holmquist (1985).

For mapping between hyper matrix spaces where $I = I(\underline{m}_u)$ and $J = I(\underline{n}_v)$ we may take the mappings i and j in the representations (5.13) be given by

$$i(k) = p^{-1}(k; \underline{m}_r) \quad \text{and} \quad j(k) = p^{-1}(k; \underline{n}_s)$$

respectively.

The main advantage of the representation of a differential given by (5.13) is the simple form of the chain-rule; if Y depends on Z which in turn depends on X , then we have

$$d_X \square Y = \left(\sum_{\ell=1}^w \partial_{y_i(\alpha)} / \partial_{z_k(\ell)} \partial_{z_k(\ell)} / \partial_{x_j(\beta)} \right)_{\substack{\alpha=1,\dots,n \\ \beta=1,\dots,m}}$$

where w is the dimension of the space of Z and k a bijective mapping from $\{1,2,\dots,w\}$ to the indexset of Z . Under the convention that $i \equiv j$ when $m=n$, this chain-rule can be written

$$d_X \square Y = (d_Z \square Y)(d_X \square Z)$$

where the juxtaposition on the right hand side denotes ordinary multiplication of matrices.

Some basic results for matrix functions

We denote by $T_{m,n}$ the direct product transposition matrix with degrees m and n , that is, the matrix which satisfies

$$(5.14) \quad \text{vec}(X^t) = T_{n,m} \text{vec}(X)$$

for a m by n matrix X . This matrix, which is a special case of the direct product permuting matrices given by Holmquist (1985), is sometimes called vec-permutation matrix, commutation matrix or permuted identity matrix, cf. Henderson and Searle (1981) for a review.

By using the relation $\text{vec}(ABC) = (A \otimes C^t)\text{vec}(B)$ for matrices A , B and C , cf. e.g. Holmquist (1985), and the decomposition rule, we obtain the product-rule

$$(5.15) \quad d_X \square (YZ) = (I \otimes Z^t)(d_X \square Y) + (Y \otimes I)(d_X \square Z)$$

for matrices Y and Z which depend on the matrix X .

We denote by $\otimes$ the half-right Kronecker product defined by Holmquist (1985) and which is related to the right Kronecker product by

$$A \otimes B = (A \otimes B)^T_{m,n}$$

for matrices A and B with m and n columns respectively.

By using the relation $\text{vec}(ABC) = (A \otimes C^t)\text{vec}(B^t)$, cf. Holmquist (1985, Theorem 3.2 (xviii)) or by using (5.14) we obtain the transposition-rule

$$d_{X^t} \square Y = (d_X \square Y)^T_{m,n}$$

for a m by n matrix X .

In the realization (5.10) of (5.1), the chain-rule has a very simple formulation; if Y depends on Z which in turn depends on X , then we have

$$(5.16) \quad d_X \square Y = (d_Z \square Y)(d_X \square Z)$$

where the juxtaposition on the right hand side denotes ordinary multiplication of matrices.

Gradients of some basic matrix functions

The proof of the following lemma illustrates the easiness of calculating differentials by using these rules based on the realization (5.10).

Expressions similar to the gradients given below can be found in papers and books by Neudecker (1968), Graybill (1969), McDonald and Swaminathan (1973), Rao (1973), Rogers (1980).

LEMMA 5.1. For arbitrary matrices A and B which do not depend on the matrix X and such that the operations are well defined, the following results hold.

- (i) $d_X \square AXB = A \otimes B^t$
- (ii) $d_X \square AX^t B = A \otimes B^t$
- (iii) $d_X \square |AXB| = |AXB| (\text{vec}(A^t (AXB)^{-t} B^t))^t$, if AXB is nonsingular
- (iv) $d_X \square \text{tr}(AXB) = (\text{vec}(A^t B^t))^t$
- (v) $d_X \square \text{tr}(AX^t B) = (\text{vec}(BA))^t$

PROOF. (i): By the product-rule we conclude that

$$\begin{aligned} d_X \square (AXB) &= (I \otimes (XB)^t) (d_X \square A) + (A \otimes I) (d_X \square (XB)) \\ &= 0 + (A \otimes I) ((I \otimes B^t) (d_X \square X) + (X \otimes I) (d_X \square B)) \\ &= (A \otimes I) (I \otimes B^t) I = A \otimes B^t. \end{aligned}$$

(ii): Let X be an m by n matrix. Then by the transposition rule

$$d_X \square AX^t B = (d_{X^t} \square AX^t B) T_{n,m} = (A \otimes B^t) T_{n,m} = A \otimes B^t.$$

(iii): Let $X = (x_{i,j})$ and denote by $c_{i,j}$ the cofactor of $x_{i,j}$. Then we have $|X| = \sum_i x_{i,\beta} c_{i,\beta}$ for any β , and hence for fixed α and β we infer that $\partial |X| / \partial x_{\alpha,\beta} = c_{\alpha,\beta}$. Thus we obtain

$$\begin{aligned} d_X \square |X| &= \sum_{\alpha,\beta} c_{\alpha,\beta} (\text{vec}(E_{\alpha,\beta}))^t = \sum_{\alpha,\beta} |X| x^{\beta,\alpha} (\text{vec}(E_{\alpha,\beta}))^t \\ &= |X| (\text{vec}(\sum_{\alpha,\beta} x^{\beta,\alpha} E_{\alpha,\beta}))^t = |X| (\text{vec} X^{-t})^t \end{aligned}$$

since $c_{\alpha,\beta} = x^{\beta,\alpha} |X|$ where $X^{-1} = (x^{i,j})$. Further by the chain-rule (5.16) we see that

$$\begin{aligned} d_X \square |AXB| &= (d_{AXB} \square |AXB|) (d_X \square AXB) \\ &= |AXB| (\text{vec}((AXB)^{-t})^t (A \otimes B^t)) = |AXB| (\text{vec}(A^t (AXB)^{-t} B^t))^t \end{aligned}$$

(iv): First we have

$$\begin{aligned} d_X \square \text{tr}(X) &= \sum_{i,j} \left(\partial \left(\sum_k x_{k,k} \right) / \partial x_{i,j} \right) (\text{vec}(E_{i,j}))^t \\ &= \sum_{i,j,k} (\partial x_{k,k} / \partial x_{i,j}) (\text{vec}(E_{i,j}))^t \\ &= \sum_k (\text{vec}(E_{k,k}))^t = (\text{vec}(I))^t. \end{aligned}$$

By the chain-rule it then follows that

$$d_X \square \text{tr}(AXB) = (\text{vec}(I))^t (A \otimes B^t) = (\text{vec}(A^t B^t))^t.$$

(v): By the chain-rule

$$d_X \square \text{tr}(AX^t B) = (\text{vec}(I))^t (A \otimes B^t) = (\text{vec}(BA))^t.$$

An alternative elegant proof of Lemma 5.1 (iii) for the special case $A = B = I$, is proposed by von Rosen (1984, p. 120) and is obtained by aid of Aitkens's integral (Aitken (1931))

$$\int_{R^n} \exp(-\frac{1}{2} x^t V x) dx = (2\pi)^{n/2} |V|^{-1/2}$$

for positive definite V . By differentiating under the integral sign we obtain

$$\begin{aligned} (5.17) \quad \int_{R^n} \exp(-\frac{1}{2} x^t V x) (-\frac{1}{2} x^t \otimes x^t) (d_V \square V) dx \\ = (2\pi)^{n/2} (-\frac{1}{2} |V|^{-3/2}) (d_V \square |V|) \end{aligned}$$

since $d_V \square x^t V x = (x^t \otimes x^t) (d_V \square V)$ by the chain-rule. Here $d_V \square V$ is not equal to the identity matrix because of the dependency between the elements of the symmetric matrix V . We will return to this subject in section 5.4. Using the vec operator we rewrite (5.17) as

$$\begin{aligned} (\text{vec}(\int_{R^n} (2\pi)^{-n/2} |V|^{\frac{1}{2}} \exp(-\frac{1}{2} x^t V x) x x^t dx))^t (d_V \square V) \\ = |V|^{-1} (d_V \square |V|) \end{aligned}$$

which by multivariate normal theory can be written

$$(\text{vec}(V^{-1}))^t (d_V \square V) = |V|^{-1} (d_V \square |V|),$$

that is

$$d_V \square |V| = |V|^{-1} (\text{vec}(V^{-1}))^t (d_V \square V).$$

We observe that the proof relies on that V is positive definite (and

hence symmetric) and the result differs by an extra factor $d_Y \square V$ from the result $d_X \square X = |X|(\text{vec}(X^{-t}))^t$ obtained from Lemma 5.1 (iii).

Gradients of inverses of matrices

By using the product-rule (5.16) we obtain the following results for generalized inverses.

THEOREM 5.3. (i) The derivative of a generalized inverse Y^- of Y satisfies the equation

$$(Y \otimes Y^t)(d_X \square Y^-) = (I \otimes I - YY^- \otimes I - I \otimes (Y^-Y)^t)(d_X \square Y).$$

(ii) The derivative of a reflexive generalized inverse Y_r^- of Y satisfies the equations

$$(Y_r^- \otimes (Y_r^-)^t)(d_X \square Y) = -(Y_r^-Y \otimes (YY_r^-)^t)(d_X \square Y_r^-)$$

and

$$(Y \otimes Y^t)(d_X \square Y_r^-) = -(YY_r^- \otimes (Y_r^-Y)^t)(d_X \square Y).$$

(iii) The derivative of the inverse Y^{-1} of Y (nonsingular) is given by

$$d_X \square Y^{-1} = -(Y^{-1} \otimes Y^{-t})(d_X \square Y).$$

PROOF. (i): Any g-inverse Y^- of Y satisfies $YY^-Y = Y$. Hence by (5.15)

$$\begin{aligned} d_X \square Y &= d_X \square YY^-Y = (I \otimes Y^t)(d_X \square YY^-) + (YY^- \otimes I)(d_X \square Y) \\ &= (I \otimes Y^t)((I \otimes (Y^-)^t)(d_X \square Y) + (Y \otimes I)(d_X \square Y^-)) \\ &\quad + (YY^- \otimes I)((d_X \square Y) \\ &= (I \otimes (Y^-Y)^t + YY^- \otimes I)(d_X \square Y) + (Y \otimes Y^t)(d_X \square Y^-) \end{aligned}$$

or

$$(I \otimes I - I \otimes (Y^-Y)^t + YY^- \otimes I)(d_X \square Y) = (Y \otimes Y^t)(d_X \square Y^-)$$

(ii): By replacing Y^- by Y_r^- in (i) and premultiplying the equation with $Y_r^- \otimes (Y_r^-)^t$ we obtain

$$(Y_r^-Y \otimes (YY_r^-)^t)(d_X \square Y_r^-) = -(Y_r^- \otimes (Y_r^-)^t)(d_X \square Y)$$

by using that $Y_r^- Y Y_r^- = Y_r^-$. By premultiplying this equation with $Y \otimes Y^t$ we obtain

$$(Y \otimes Y^t)(d_X \square Y_r^-) = -(Y Y_r^- \otimes (Y_r^- Y)^t)(d_X \square Y)$$

by using that $Y Y_r^- Y = Y$.

(iii): The result immediately follows from the first equation of (ii) by replacing Y_r^- by Y^{-1} . $\square$

To obtain explicit expressions for gradients of generalized inverses we need the following lemma.

LEMMA 5.2. A necessary and sufficient condition for G to satisfy

$$\begin{aligned} (Y Y^+ \otimes (Y^+ Y)^t - Y Y^+ \otimes (Y^- Y)^t - Y Y^- \otimes (Y^+ Y)^t)G \\ = (I \otimes I - Y Y^- \otimes I - I \otimes (Y^- Y)^t)G \end{aligned}$$

for an arbitrary g -inverse Y^- and the Moore-Penrose inverse Y^+ is that

$$G = (Y Y^+ \otimes (Y^+ Y)^t + (I - Y Y^+) \otimes (Y^- Y)^t + Y Y^- \otimes (I - (Y^+ Y)^t))H$$

for arbitrary H .

PROOF. Rewriting the first equation in the lemma, we have

$$(I \otimes I - Y Y^+ \otimes (Y^+ Y)^t - (I - Y Y^+) \otimes (Y^- Y)^t - Y Y^- \otimes (I - (Y^+ Y)^t))G = 0.$$

Write $P = I \otimes I - Y Y^+ \otimes (Y^+ Y)^t - (I - Y Y^+) \otimes (Y^- Y)^t - Y Y^- \otimes (I - (Y^+ Y)^t)$.

It is easily verified that P is idempotent. Writing $G = PG + (I-P)G$

we see that $PG = 0$ implies that $G = (I-P)G$ which in turn implies that

$G = (I-P)H$ for some H . On the other hand if $G = (I-P)H$ then $PG = 0$. $\square$

THEOREM 5.4. A necessary and sufficient condition for $d_X \square Y^-$ to exist is

that $d_X \square Y = (Y Y^+ \otimes (Y^+ Y)^t + (I - Y Y^+) \otimes (Y^- Y)^t + Y Y^- \otimes (I - (Y^+ Y)^t))H$

for some H . In this case

$$d_X \square Y^- = -(Y^+ \otimes (Y^+)^t)(d_X \square Y) + (I \otimes I - Y^+ Y \otimes (Y Y^+)^t)K$$

for some K .

PROOF. A necessary and sufficient condition for the equation

$$(Y \otimes Y^t)(d_X \square Y^-) = (I \otimes I - YY^- \otimes I - I \otimes (Y^-Y)^t)(d_X \square Y)$$

to have a solution $d_X \square Y^-$ is that

$$\begin{aligned} (Y \otimes Y^t)(Y^+ \otimes (Y^t)^+)(I \otimes I - YY^- \otimes I - I \otimes (Y^-Y)^t)(d_X \square Y) \\ = (I \otimes I - YY^- \otimes I - I \otimes (Y^-Y)^t)(d_X \square Y) \end{aligned}$$

see e.g. Rao and Mitra (1971, Theorem 2.3.2) which by Lemma 5.2 is satisfied if and only if $d_X \square Y^-$ is of the form specified in the theorem. By Theorem 2.3.2 of Rao and Mitra (1971) it follows that the solution in this case is given by

$$\begin{aligned} d_X \square Y^- &= (Y^+ \otimes (Y^t)^+)(I \otimes I - YY^- \otimes I - I \otimes (Y^-Y)^t)(d_X \square Y) \\ &\quad + (I \otimes I - Y^+Y \otimes (YY^+)^t)K \\ &= (Y^+ \otimes (Y^+)^t)(YY^+ \otimes (Y^+Y)^t + (I - YY^+) \otimes (Y^-Y)^t \\ &\quad + YY^- \otimes (I - (Y^+Y)^t))(d_X \square Y) \\ &\quad + (I \otimes I - Y^+Y \otimes (YY^+)^t)K \\ &= (Y \otimes (Y^+)^t)(d_X \square Y) + (I \otimes I - Y^+Y \otimes (YY^+)^t)K. \end{aligned}$$

This proves the theorem. □

5.4. Symmetric matrix derivatives

Problems arise when the technique of using differential operators is applied to situations where elements of a space are expressed in terms of a basis of a larger space, as for example when symmetric matrices are expressed in terms of the canonical basis for square matrices.

If f is a mapping from the space S_+^m of symmetric m -square matrices to $K^{r,s}$, then the differential of f at $X \in S_+^m$ can not be calculated as $d_X \square f(X)$ or $D_X \otimes f(X)$, as given in subsection 5.3, without taking special considerations into account, since d_X and D_X are not differential operators in the space of m -square symmetric matrices but in

the space of m -square matrices.

There are in the literature two different approaches for treating this problem.

(i) One technique considered by McCulloch (1982) is to map the representation into another representation space of proper dimension in which a differential operator is defined. McCulloch uses the vech ("vec-half") operator to map the representation as a m -square symmetric matrix into a column vector consisting of the $m(m+1)/2$ distinct coordinates. The unique matrix $G \in \mathbb{R}^{m^2, m(m+1)/2}$ such that

$$\text{vec}(X) = G\text{vech}(X)$$

is then used to obtain an expression for

$$d_h(X) \square f(X) \equiv \text{vec}(f(X))(\text{vech}(\partial/\partial X))^t$$

in terms of $d_{\{X\}} \square f(X)$ which is given by

$$(5.18) \quad d_h(X) \square f(X) = (d_{\{X\}} \square f(X))G.$$

This relation follows from the chain-rule.

Nel (1980) considers more general patterned matrices, that is matrices with functionally dependent elements. If X is a m by n matrix with q mathematically independent elements, s repeated or functionally dependent elements and v constant elements ($q+s+v=mn$) we denote by $\text{vecp}(X)$ (p for patterned) the column vector obtained from X in a manner similar to how $\text{vec}(X)$ is obtained, but including only the q mathematically independent elements. If the s dependent elements are linearly dependent of the q independent elements then there exist a mn by q matrix G such that

$$(5.19) \quad \text{vec}(X) = G\text{vecp}(X).$$

From this expression we can obtain an expression for

$$d_p(X) \square f(X) \equiv \text{vec}(f(X))(\text{vecp}(\partial/\partial X))^t$$

in terms of $d_X \square f(X)$, namely

$$(5.20) \quad d_{p(X)} \square f(X) = (d_{\{X\}} \square f(X))G.$$

This follows from the chain-rule.

The relation (5.20) is in general true even if the dependence is non-linear. In this case (5.19) is not satisfied, but if G is defined by

$$(5.21) \quad G \equiv d_{p(X)} \square X = \text{vec}(X)(\text{vecp}(\partial/\partial X))^t$$

then (5.20) holds; this follows by the chain-rule. In the special case when the elements are linearly dependent, G given by (5.21) also satisfies (5.19).

(ii) Another method is to relate the differential operator in S_+^m to differential operators in $K^{m,m}$. This approach is considered by Rogers (1980, ch. 10) and can be exemplified by the case when f is scalar valued, in which case Rogers gives the relation

$$(5.22) \quad D_X \otimes f(X) = D_{\{X\}} \otimes f(X) + D_{\{X^t\}} \otimes f(X) - \text{Diag}(D_{\{X\}} \otimes f(X))$$

where $\text{Diag}(X)$ is the diagonal matrix having the same diagonal elements as X . Here the notation $\{X\}$ on the right hand side indicates that the differential should be calculated as if there were no restriction on the elements of X . Such a relation has been used by other authors to obtain derivatives with symmetric X ; cf. e.g. Rao (1973, p. 72). The relation (5.22) is obtained by componentwise differentiation but the interpretation of the expression as a vector space operation is not clear.

The relation (5.22) is connected with symmetric X . When the dependency between the elements of X is of another type, other relations has to be developed.

We note that (5.22) can be written

$$(5.23) \quad d_X \square f(X) = (d_{\{X\}} \square f(X))H$$

by aid of the vec operator; here $H = d_X \square X = d_{\{X\}} \square X + d_{\{X^t\}} \square X - d_{\{X\}} \square \text{Diag}(X)$. This relation does not necessarily restrict to scalar

valued functions f and is of a form similar to (5.20). However, while the dimension of the covariant basis of (5.20) is equal to the number of independent variables $(m(m+1)/2)$, the dimension of the covariant basis of (5.23) is equal to the number of variables of the representation (m^2) .

As we shall see in Example 5.4 below, such a definition of a differential mapping causes the mapping to act very strangely on its domain space (here S_+^m) in order to be the true differential.

EXAMPLE 5.4. Let $X \in S_+^m$ be expressed in the canonical basis for $K^{m,m}$ as

$$X = \sum_{i < j} x_{ij} (E_{ij} + E_{ji}) + \sum_{i=1}^m x_{ii} E_{ii}$$

and in an orthonormal basis $\{F_{ij}, i \leq j\}$ for S_+^m as

$$X = \sum_{i \leq j} x'_{ij} F_{ij}.$$

For example we can take

$$F_{ij} = \begin{cases} (E_{ij} + E_{ji})/\sqrt{2} & i < j \\ E_{ii} & i = j \end{cases}$$

in which case

$$x'_{ij} = \begin{cases} \sqrt{2} x_{ij} & i < j \\ x_{ii} & i = j \end{cases}$$

For simplicity take $m=2$ and let $Y = f(X) = x_{11} + x_{12}^2 = x'_{11} + x_{12}'^2/2$.

The technique of Rogers and Rao gives

$$D_X \otimes Y = \begin{pmatrix} 1 & 0 \\ 2x_{12} & 0 \end{pmatrix} + \begin{pmatrix} 1 & 2x_{12} \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 2x_{12} \\ 2x_{12} & 0 \end{pmatrix}.$$

and

$$d_X \square Y = (1, 2x_{12}, 2x_{12}, 0).$$

The technique of McCulloch and Nel gives

$$d_h(X) \square Y = (1, 2x_{12}, 0).$$

For $Z = \begin{pmatrix} z_{11} & z_{12} \\ z_{12} & z_{22} \end{pmatrix} \in S_+^2$, we have

$$f(X + Z) - f(X) = z_{11} + z_{12}^2 + 2z_{12}x_{12} = z'_{11} + z'_{12}^2/2 + z'_{12}x'_{12}.$$

Thus a direct application of (3.1) yields that the differential of f at $X \in S_+^2$ in direction Z is

$$(5.24) \quad z_{11} + 2z_{12}x_{12} = z'_{11} + z'_{12}x'_{12}.$$

By use of (5.22) and (5.23) this can be written

$$(5.25) \quad \text{tr}((D_X \otimes Y)(Z + \text{Diag}(Z))/2) = \text{tr}\left(\begin{pmatrix} 1 & 2x_{12} \\ 2x_{12} & 0 \end{pmatrix} \begin{pmatrix} z_{11} & \frac{1}{2}z_{12} \\ \frac{1}{2}z_{12} & z_{22} \end{pmatrix}\right)$$

and

$$(5.26) \quad (d_X \square Y)\text{vec}((Z + \text{Diag}(Z))/2).$$

Here the somewhat strange argument $(Z + \text{Diag}(Z))/2$ is caused by the subtraction of $\text{Diag}(D_{\{X\}} \otimes f(X))$ in formula (5.22) and $d_{\{X\}} \square \text{Diag}(X)$ in (5.23).

By use of (5.18) the differential (5.24) can be written

$$(d_{h(X)} \square Y)\text{vech}(Z)$$

which uses an inner product in S_+^m instead of in $\mathbb{R}^{m,m}$. □

The question arises how to define the differential so that the value at Z of the differential of f at X can be written

$$(d_X \square Y)\text{vec}(Z)$$

as in the case when all elements are mathematically independent, that is which uses the inner product in the space in which the argument is represented.

In order to obtain results for differential with respect to symmetric matrices we shall use the results for differentials of extended mappings and their relation to the differentials of the original mappings. We regard $K^{m,m}$ as a direct sum of S_+^m and S_-^m , where S_-^m is the space of

skew-symmetric matrices. For a mapping f defined on S_+^m we define a mapping f' , with domain $K^{m,m} = S_+^m \oplus S_-^m$, by

$$f'(X) = f(X_+)$$

for every $X = X_+ + X_- \in K^{m,m}$, where $X_+ \in S_+^m$ and $X_- \in S_-^m$. The representation of f' is given as follows: if $f(X_+)$ is a matrix expression depending on the m -square symmetric matrix X_+ , then $f'(X)$ is the same expression but with X_+ replaced by an m -square matrix X . By (3.4) we have that $D_{S_+^m} f'$ and $Df \circ P_{S_+^m}$ coincide on S_+^m for every $X \in K^{m,m}$ at which the differential is taken, that is for every $X = X_+ + X_- \in K^{m,m}$ and $Z_+ \in S_+^m$ we have

$$D_{S_+^m} f'(X)(Z_+) = Df(X_+)(Z_+).$$

The mapping $Df'(X)$ can be realized as $d_X \square f'(X)$, which is equivalent to $d_{\{X_+\}} \square f(X_+)$. Hence $D_{S_+^m} f'(X)$ can be realized as $(d_X \square f'(X))S$ where $S = \frac{1}{2}(d_X \square X + d_{X^t} \square X) = \frac{1}{2}(I_m \otimes I_m + T_{m,m})$ is the projection operator on S_+^m . We thus have that

$$(5.27) \quad d_{X_+} \square f(X_+) = (d_{\{X_+\}} \square f(X_+))S = \frac{1}{2}(d_{\{X_+\}} \square f(X_+) + d_{\{X_+^t\}} \square f(X_+)).$$

EXAMPLE 5.4. (Continued) An application of (5.27) yields that the differential is

$$\text{tr}(\frac{1}{2}(D_{\{X\}} \otimes Y + D_{\{X^t\}} \otimes Y)Z) = \text{tr}\left(\begin{pmatrix} 1 & x_{12} \\ x_{12} & 0 \end{pmatrix} \begin{pmatrix} z_{11} & z_{12} \\ z_{12} & z_{22} \end{pmatrix}\right)$$

which is equal to (5.24) as it should.

EXAMPLE 5.5. A well-known relation for the multinormal density

$p = p(x, \mu, \Sigma) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp(-\frac{1}{2}(x-\mu)^t \Sigma^{-1} (x-\mu))$ is often stated elementwise as

$$(5.28a) \quad \frac{\partial p}{\partial \sigma_{ij}} = \frac{\partial^2 p}{\partial x_i \partial x_j} \quad \text{if } i \neq j$$

and

$$(5.28b) \quad \frac{\partial p}{\partial \sigma_{ii}} = \frac{1}{2} \frac{\partial^2 p}{\partial x_i^2}$$

where $x^t = (x_1, \dots, x_n)$ and $\Sigma = (\sigma_{ij})$. The result (5.28a) is stated by Plackett (1954) without specifying that $i \neq j$, while Patil and Boswell (1970) state the results (5.28a,b) and also shows that these relations characterize the normal density. In matrix notation these relations can be written

$$\begin{aligned} D_{\Sigma} \otimes p(x, \mu, \Sigma) &= D_x^2 \otimes p(x, \mu, \Sigma) - \frac{1}{2} \text{Diag}(D_x^2 \otimes p(x, \mu, \Sigma)) \\ &= \frac{1}{2} (D_x^2 \otimes p(x, \mu, \Sigma) + D_x^2 \otimes p(x, \mu, \Sigma) - \text{Diag}(D_x^2 \otimes p(x, \mu, \Sigma))) \end{aligned}$$

where $D_x^2 = (\partial^2 / \partial x_i \partial x_j)$, or as

$$d_{\Sigma} \square p(x, \mu, \Sigma) = \frac{1}{2} (\text{vec}((d_x \square (d_x \square p(x, \mu, \Sigma))))^t H$$

where H is given below (5.23). The relations (5.28a,b) are obtained by elementwise differentiation, regarding σ_{ij} and σ_{ji} as the same element. The interpretation of these relations is however only of interest in connection with their operation on their arguments. Analogously to (5.25) and (5.26) we have that for a symmetric matrix Z the following relation holds.

$$\begin{aligned} (d_{\Sigma} \square p(x, \mu, \Sigma)) \text{vec}(\frac{1}{2}(Z + \text{Diag}(Z))) \\ = \text{tr}(\frac{1}{2}(d_x \square (d_x \square p(x, \mu, \Sigma))) H (\frac{1}{2}(Z + \text{Diag}(Z)))). \end{aligned}$$

However, we can analogously to (5.27) write the relation

$$\begin{aligned} (5.29) \quad d_{\Sigma} \square p(x, \mu, \Sigma) &= \frac{1}{2} (\text{vec}((d_x \square (d_x \square p(x, \mu, \Sigma))))^t S \\ &= \frac{1}{2} (\text{vec}((d_x \square (d_x \square p(x, \mu, \Sigma))))^t \end{aligned}$$

where $S = \frac{1}{2}(I_n \otimes I_n + T_{n,n})$ and where the last equality follows from that $d_x \square (d_x \square p(x, \mu, \Sigma))$ is symmetric.

The mappings in the relation (5.29) act on their arguments as

$$(d_{\Sigma} \square p(x, \mu, \Sigma)) \text{vec}(Z) = \text{tr}(\frac{1}{2}(d_x \square (d_x \square p(x, \mu, \Sigma))) Z)).$$

We note that (5.29) can be interpreted as the Kolmogorov backward equation for the standard n -dimensional Brownian motion where the "time" parameter here is Σ so that it is a n -dimensional Wiener field on S_+^n .

Following this idea for the multivariate normal distribution with mean depending on the covariance, we have the following result.

THEOREM 5.5. Let $p(x, \mu(\Sigma), \Sigma) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp(-\frac{1}{2}(x-\mu(\Sigma))^t \Sigma^{-1} (x-\mu(\Sigma)))$ be the multivariate normal density with expectation $\mu(\Sigma)$ depending on the covariance matrix Σ . Then

$$(5.30) \quad d_\Sigma \square p(x, \mu(\Sigma), \Sigma) = \frac{1}{2} (\text{vec}(d_x \square (d_x \square p(x, \mu(\Sigma), \Sigma))))^t + (d_x \square p(x, \mu(\Sigma), \Sigma)) (d_\Sigma \square \mu(\Sigma)).$$

Especially if $\mu(\Sigma) = B \text{vec}(\Sigma) + \alpha$, it holds that

$$d_\Sigma \square p(x, \mu(\Sigma), \Sigma) = \frac{1}{2} (\text{vec}(d_x \square (d_x \square p(x, \mu(\Sigma), \Sigma))))^t + (d_x \square p(x, \mu(\Sigma), \Sigma)) B \frac{1}{2} (I_n \otimes I_n + T_{n,n})$$

and if $\mu(\Sigma) = \Gamma \Sigma \beta + \alpha$, then

$$d_\Sigma \square p(x, \mu(\Sigma), \Sigma) = \frac{1}{2} (\text{vec}(d_x \square (d_x \square p(x, \mu(\Sigma), \Sigma))))^t + (d_x \square p(x, \mu(\Sigma), \Sigma)) (\Gamma \otimes \beta^t) \frac{1}{2} (I_n \otimes I_n + T_{n,n}).$$

PROOF. We have from Lemma 5.1 (i) and (ii) and the productrule that

$$d_x \square (x - \mu(\Sigma))^t \Sigma^{-1} (x - \mu(\Sigma)) = 2(x - \mu(\Sigma))^t \Sigma^{-1}.$$

Thus by the chain-rule

$$d_x \square p(x, \mu(\Sigma), \Sigma) = -p(x, \mu(\Sigma), \Sigma) (x - \mu(\Sigma))^t \Sigma^{-1}$$

and by the product-rule

$$\begin{aligned} d_x \square (d_x \square p(x, \mu(\Sigma), \Sigma)) &= \Sigma^{-1} (x - \mu(\Sigma)) (x - \mu(\Sigma))^t \Sigma^{-1} p(x, \mu(\Sigma), \Sigma) \\ &\quad - p(x, \mu(\Sigma), \Sigma) \Sigma^{-1} \\ &= \Sigma^{-1} (x - \mu(\Sigma)) (x - \mu(\Sigma))^t \Sigma^{-1} - \Sigma^{-1} p(x, \mu(\Sigma), \Sigma). \end{aligned}$$

From the product-rule and the results of Lemma 5.1 we have that

$$d_\Sigma \square p(x, \mu(\Sigma), \Sigma) = (p(x, \mu(\Sigma), \Sigma) \frac{1}{2} (\text{vec}(\Sigma))^t (d_{\{\Sigma\}} \square \Sigma^{-1}) S$$

$$\begin{aligned}
 & - \frac{1}{2} p(x, \mu(\Sigma), \Sigma) ((x - \mu(\Sigma))^{\langle 2 \rangle t} - 2(x - \mu(\Sigma))^t \Sigma (d_{\{\Sigma\}^{-1}} \square \mu(\Sigma))) \\
 & \quad (d_{\{\Sigma\}} \square \Sigma^{-1}) S \\
 & = -\frac{1}{2} p(x, \mu(\Sigma), \Sigma) ((\text{vec}(\Sigma^{-1}))^t - ((x - \mu(\Sigma))^t \Sigma^{-1})^{\langle 2 \rangle t}) S \\
 & + p(x, \mu(\Sigma), \Sigma) (x - \mu(\Sigma))^t \Sigma^{-1} (d_{\{\Sigma\}} \square \mu(\Sigma)) S \\
 & = -\frac{1}{2} p(x, \mu(\Sigma), \Sigma) (\text{vec}(\Sigma^{-1} - \Sigma^{-1} (x - \mu(\Sigma)) (x - \mu(\Sigma))^t \Sigma^{-1}))^t \\
 & + p(x, \mu(\Sigma), \Sigma) (x - \mu(\Sigma))^t \Sigma^{-1} (d_{\{\Sigma\}} \square \mu(\Sigma)) S
 \end{aligned}$$

From these expressions we infer that

$$\begin{aligned}
 d_{\Sigma} \square p(x, \mu(\Sigma), \Sigma) & = \frac{1}{2} (\text{vec}(d_x \square (d_x \square p(x, \mu(\Sigma), \Sigma)))^t \\
 & + (d_x \square p(x, \mu(\Sigma), \Sigma)) (d_{\Sigma} \square \mu(\Sigma))
 \end{aligned}$$

which proves the first part.

Especially $d_{\Sigma} \square (B\text{vec}(\Sigma) + \alpha) = BS$ as follows from Lemma 5.1 (i), which proves the second part. Since $d_{\Sigma} \square \Gamma \Sigma \beta = (\Gamma \otimes \beta^t) S$, as follows from Lemma 5.1 (i), the third part follows which completes the proof of the theorem. $\square$

We note that the relation (5.30) is the Kolmogorov backward equation for a Brownian motion with drift.

5.5. Matrix derivatives for matrices with linear structure

The result (5.27) for symmetric matrices can be generalized to matrix derivatives with respect to matrices with linear dependencies between the elements. The problem by using different representations is the preservation of the value of the gradient in a point and in a certain direction. The question how to define differential operators so that they can be applied to mappings on subspaces can be answered by means of partial isometries. Let $\langle \cdot, \cdot \rangle_X$ and $\langle \cdot, \cdot \rangle_Y$ be inner products in X and Y respectively. Let $A \in L(X; Y)$ and $\partial/\partial x$ be a differential operator in X at x . We define a differential operator $\partial/\partial y$ at $y = Ax$ in Y such that

$$(5.31) \quad \langle (\partial/\partial Ax)^*, Au \rangle_Y = \langle (\partial/\partial x)^*, u \rangle_X$$

for all $u, x \in M(A^*)$. Here $M(A^*)$ is the subspace in X which is orthogonal to the null space of A . The relation (5.31) is satisfied if A is a partial isometry, cf. e.g. Rao and Mitra (1971, Sec. 5.7). This is equivalent to that A^* is the Moore-Penrose inverse of A and hence is A^*A the projection operator onto $M(A^*)$ and AA^* is the projection operator onto $M(A)$. The relation (5.31) can thus be written

$$(5.32) \quad \langle (AA^*)^* (\partial/\partial Ax)^*, Au \rangle_Y = \langle (\partial/\partial x)^*, u \rangle_X$$

since $\langle (AA^*)^* (\partial/\partial Ax)^*, Au \rangle_Y = \langle (\partial/\partial Ax)^*, AA^* Au \rangle_Y = \langle (\partial/\partial Ax)^*, Au \rangle_Y$.

As a consequence of relation (5.32) we see that if A is an orthogonal projection operator, $AA^* = A$ and a differential operator on the subspace Y of X is given by $(\partial/\partial x)A$ that is

$$(d_x \square f(x))A = d_y \square f(y).$$

We look more closely at two particular cases.

Explicitly given subspace

Let $y - b \in M(B)$, where $M(B)$ is the range space of $B \in L(X; Y)$, and $b \in Y$, that is y can be written $y = Bx + b$ for some $x \in X$. We shall assume that B has full column rank, that is there exist a mapping $B_L^{-1} \in L(Y; X)$ such that $B_L^{-1}B = I_X$. Here B_L^{-1} is not generally unique.

We immediately obtain from Theorem 3.1 (ii) that

$$d_x \square y = B.$$

The relation $y = Bx + b$ implies that for a given value of y , the possible values of x is given by

$$x = B_L^{-1}(y - b)$$

for all possible left inverses B_L^{-1} . For each of these left inverses

$$d_y \square x = B_L^{-1}.$$

By the chainrule we have

$$d_x \square x = (d_y \square x)(d_x \square y) = B_L^{-1}B = I_X,$$

while

$$d_y \square y = (d_x \square y)(d_y \square x) = BB_L^{-1}.$$

Here BB_L^{-1} is the projection operator onto $M(B)$. Any left inverse B_L^{-1} can be written $B_L^{-1} = (\bar{B}^tVB)^{-1}\bar{B}^tV$ for arbitrary V such that $\text{rank}(\bar{B}^tVB) = \text{rank}(B)$, where $\bar{B}^t$ is the conjugate dual to B (cf. Rao and Mitra, 1971). However, $B_L^{-1}B$ is invariant for any choice of V .

The transition matrix K_{pq} of Nel (1980), defined for the particular manifolds on $R^{n,n}$ considered there, is a particular choice of B_L^{-1} . The definition is chosen so that $B_L^{-1} = B^+$, the Moore-Penrose inverse of B .

For inner product spaces we can also write $B_L^{-1} = (B^*B)^{-1}B^*$ where the adjoint $B^* = W^{-1}\bar{B}^tV$ for positive definite linear mappings $W \in GL(X)$ and $V \in GL(Y)$. Here BB_L^{-1} is the orthogonal projection onto $M(B)$.

An example of this case is given by (5.18). In the special case when vecp is the vech operator, we have

$$G(G^tG)^{-1}G^t = S = \frac{1}{2}(I_n \otimes I_n + T_{n,n}).$$

In the case B does not have full column rank, the projection operator onto $M(B)$ can be written $B(\bar{B}^tVB)^{-1}\bar{B}^tV$ for any generalized inverse.

Implicitely given subspace

Let the subspace for y be given by $Ay = b$ for specified $A \in L(Y;X)$ where A has full row rank and $b \in X$. This restriction implies that the possible values of y are given by

$$y = A_R^{-1}b + (I_Y - A_R^{-1}A)z$$

for arbitrary $z \in Y$ and some specified right inverse A_R^{-1} to A . We have that

$$d_z \square y = (I_Y - A_R^{-1}A).$$

The projection operator on the linear part of the affine range space of y is $I_Y - A_R^{-1}A$. The chosen right inverse can be written $A_R^{-1} = W\bar{A}^t(AW\bar{A}^t)^{-1}$ for arbitrary W such that $\text{rank}(AW\bar{A}^t) = \text{rank}(A)$. Alternatively we write $A_R^{-1} = A^*(AA^*)^{-1}$ where the adjoint $A^* = W\bar{A}^tV^{-1}$ is taken with respect to the p.d. matrices $W \in GL(Y)$ and $V \in GL(X)$. The matrix $I_Y - A_R^{-1}A$ is idempotent of rank $\dim Y - \text{rank} A$.

Again an example of this case is afforded by the case of symmetric matrices. For symmetric n -square Σ there exist a $\frac{1}{2}n(n-1)$ by n^2 matrix A of rank $\frac{1}{2}n(n-1)$ such that

$$\text{Avec} \Sigma = 0.$$

For this matrix A ,

$$I_{n^2} - A_R^{-1}A = S = \frac{1}{2}(I_n \otimes I_n + T_{n,n})$$

which has rank $n^2 - \frac{1}{2}n(n-1) = \frac{1}{2}n(n+1)$.

6. COORDINATE SPACE MOMENTS AND CUMULANTS: MOMENT AND CUMULANT ARRAYS

If x is a random column vector of coordinates with respect to the canonical-basis of K^m , then the mean $\mu_{1,x} = E(x)$ of x is usually represented as a column vector of expectations of the coordinates of x . The second order central moment is usually represented as a dispersion matrix $E((x-E(x))(x-E(x))^T)$. We shall use the different but isomorphic representation as

$$E((x-E(x))^{<2>}) \equiv E((x-E(x)) \otimes (x-E(x))).$$

Here $\otimes$ denotes the Kronecker product as defined in Example 5.2. Such a representation of the dispersion can easily be generalized to moments of the random vector of arbitrary order. It has also the advantage that moments of random matrices can be defined without making a preparatory transformation into a vector as for example when the dispersion matrix of a random matrix X is defined as being the dispersion matrix of $\text{vec}(X)$. For more general array representations of random variables the moments can be represented as a corresponding (generalized) Kronecker power of these representations. The r 'th moment of a random hypermatrix $X = (x_{\underline{i}_k})$ in k dimensions can be represented as $E(X^{<r>})$ where for two hypermatrices A and $B = (b_{\underline{i}_k})$ the (generalized) Kronecker product $A \otimes B$ is defined by $A \otimes B = (ab_{\underline{i}_k})$. The corresponding r 'th central moment is realized as $E((X-E(X))^{<r>})$.

With a specific realization of a moments is connected a specific realization of the r 'th derivative of the m.g.f. When the r 'th moment is represented as $E(X^{<r>})$, the derivatives of the m.g.f. should be represented as $D_T^{<r>} \otimes M_X(T)|_{T=0}$ which gives the relation

$$(6.1) \quad D_T^{<r>} \otimes M_X(T)|_{T=0} = (E(X^{<r>}))^*.$$

Generally we have for matrices X that $X^{<r>*} = X^{*<r>}$ and thus for self-adjoint matrices X the moments are obtained directly from the

differentials of the m.g.f.

In case T has restrictions on its elements, e.g. being symmetric, the derivative in (6.1) has to be calculated with taking this into account as described in section 5. However it is simpler to use the result of Theorem 4.7 and calculate the derivatives of the extended moment generating function. The projection operator on S_+^m along S_-^m of m -square matrices is given by $S = \frac{1}{2}(I_m \otimes I_m + T_{m,m})$ which acts on $Y \in K^{m,m}$ as $\text{Svec}(Y) = \text{vec}(\frac{1}{2}(Y+Y^t))$. For m.g.f.'s of X expressed in symmetric T , the r 'th moment adjoint can be obtained as

$$D_{\{T\}}^{\langle r \rangle} \otimes M_X(\frac{1}{2}(T+T^t))|_{T=0} = (E(X^{\langle r \rangle}))^*.$$

Similarly the r 'th cumulant adjoint can be obtained as

$$(6.2) \quad D_{\{T\}}^{\langle r \rangle} \otimes (\ln M_X(\frac{1}{2}(T+T^t)))|_{T=0} = (E(X^{\langle r \rangle}))^*.$$

Here the differentiations immediately give the moment and cumulant adjoints since the difference between these adjoints and the corresponding mappings is a matter of convention.

The relations between moments and cumulants given in a coordinate free setting in section 4, have corresponding versions in coordinate spaces. Such relations requires properly defined symmetrizers on coordinate spaces. In order to define these on R^n , $R^{m,n}$ and more generally on $R^{\frac{n}{r}}$, the space of $\underline{n}_r$ -hypermatrices, we need some further notations.

Let $\sigma \in S_{\underline{1}_r}$ be given by (2.4). We define the action $\sigma: I(\underline{n}_r) \rightarrow \sigma I(\underline{n}_r)$ by

$$\sigma \underline{i}_r = (i_{\sigma^{-1}(1)}, \dots, i_{\sigma^{-1}(r)}) \in \sigma I(\underline{n}_r) = I(\sigma \underline{n}_r)$$

for $\underline{i}_r \in I(\underline{n}_r)$.

The direct product permuting matrix (DPPM) with r degrees n and permutation argument $\sigma \in S_{\underline{1}_r}$,

$$P_{\underline{n}_r}(\sigma) = \sum_{\underline{i}_r \in I(\underline{n}_r)} E_{P(\sigma \underline{i}_r; \underline{n}_r), P(\underline{i}_r; \underline{n}_r)}$$

is the permutation operator on $(\mathbb{R}^n)^{<r>}$ associated with σ , that is,

$$P_{n \atop -r}(\sigma)(x_1 \otimes \dots \otimes x_r) = x_{\sigma^{-1}(1)} \otimes \dots \otimes x_{\sigma^{-1}(r)}$$

where $x_1, \dots, x_r \in \mathbb{R}^n$. The permutation operator on $(\mathbb{R}^{m,n})^{<r>}$ associated with σ is $P_{m \atop -r}(\sigma) \otimes P_{n \atop -r}(\sigma)$, that is,

$$\begin{aligned} (P_{m \atop -r}(\sigma) \otimes P_{n \atop -r}(\sigma)) \text{vec}(A_1 \otimes \dots \otimes A_r) \\ = \text{vec}(A_{\sigma^{-1}(1)} \otimes \dots \otimes A_{\sigma^{-1}(r)}) \end{aligned}$$

or equivalent

$$(6.3) \quad P_{m \atop -r}(\sigma)(A_1 \otimes \dots \otimes A_r)P_{n \atop -r}(\sigma)^t = A_{\sigma^{-1}(1)} \otimes \dots \otimes A_{\sigma^{-1}(r)}$$

where $A_1, \dots, A_r \in \mathbb{R}^{m,n}$; cf. Holmquist (1985) where a generalization is given for hypermatrix spaces. The symmetrizer $S_{\underline{r}_k}$ corresponding to a partition $\underline{r}_k$ of weight k of r is then defined in analogy with the coordinate free version: On $(\mathbb{R}^n)^{<r>}$ the symmetrizer is given by

$$S_{\underline{r}_k} = \sum_{\sigma \in S_{\underline{r}_k}} P_{n \atop -r}(\sigma) / \left[\begin{matrix} r \\ \underline{r}_k \end{matrix} \right]$$

and on $(\mathbb{R}^{m,n})^{<r>}$ the symmetrizer acts on its argument as

$$S_{\underline{r}_k} \text{vec}(A)$$

where $A \in \mathbb{R}^{mr, nr}$ and

$$S_{\underline{r}_k} = \sum_{\sigma \in S_{\underline{r}_k}} P_{m \atop -r}(\sigma) \otimes P_{n \atop -r}(\sigma) / \left[\begin{matrix} r \\ \underline{r}_k \end{matrix} \right].$$

7. THE CUMULANTS AND MOMENTS OF THE NON-CENTRAL WISHART DISTRIBUTION OVER REAL SYMMETRIC MATRICES

7.1. Introduction

In this section we give an example of the technique for deriving moments and cumulants by differentiating generating functions. We shall derive the moments about zero and the cumulants, in matrix formulation, of the q -square real symmetric random matrix X whose m.g.f. at T is given by

$$(7.1) \quad M_X(T) = |I - BT|^{-\alpha} \exp(\text{tr}(-\Lambda T(I - BT)^{-1}))$$

for symmetric T such that $(I - BT)$ is non-singular, where B and Λ are q -square positive semi-definite (p.s.d.) matrices and α is a real scalar larger than $(q-1)/2$ or equal to $1/2, 1, 3/2, \dots, (q-1)/2$. For $B = 2V$ and $\alpha = n/2$, this is the m.g.f. of $W_q(n, V, \Lambda)$, the non-central Wishart distribution with n degrees of freedom and noncentrality parameter Λ . The first moments of this distribution are given element-wise by Anderson (1958, p. 161) in the central case ($\Lambda = 0$). The two first moments were given in matrix formulation by Magnus and Neudecker (1979) for general Λ . These were derived by using the fact that X has the same distribution as $\sum_{i=1}^n x_i x_i^t$, for $x_i \in N_q(m_i, V)$, x_i and x_j , being independent for $i \neq j$ and where the non-centrality parameter Λ is equal to $\sum_{i=1}^n m_i m_i^t$. De Waal and Nel (1973) derive the expectations of X , X^2 , X^3 and X^4 in the central case by elementwise differentiation of the m.g.f. and Neudecker and Wansbeek (1983) give the expectation of $X^{<4>}$ in the central case.

7.2. Notations

Let $\underline{w}_{r-1} = (w_1, \dots, w_{r-1})$ be a binary $(r-1)$ -tuple of 1's and t 's and denote by $\{1, t\}^{r-1}$ the set of all such $(r-1)$ -tuples. We define for $\sigma \in S_{1-r-1}$ and $\underline{w}_{r-1} \in \{1, t\}^{r-1}$,

$$k' = k'(\sigma, \underline{w}_{r-1}) = \begin{cases} \sigma^{-1}(k) + 1_{\{w_{\sigma^{-1}(k)} = 1\}} & k=1,2,\dots,r-1 \\ r & k=r \end{cases}$$

and

$$k'' = k''(\sigma, \underline{w}_{r-1}) = \begin{cases} \sigma^{-1}(k) + 1_{\{w_{\sigma^{-1}(k)} = t\}} & k=1,2,\dots,r-1 \\ 1 & k=r. \end{cases}$$

The pre-gamma matrix associated with $\sigma \in S_{\underline{1}_{r-1}}$, with r degrees q and weight $\underline{w}_{r-1} \in \{1,t\}^{r-1}$, is defined by

$$G_{q\underline{1}_r}(\sigma; \underline{w}_{r-1}) = \sum_{\underline{i}_r \in I(q\underline{1}_r)} E_{p(i_1'', \dots, i_r''), p(i_1, \dots, i_r)}.$$

We also define, for q -square matrices $A_1, \dots, A_r$,

$$(7.2) \quad H_{q\underline{1}_r}(A_1, \dots, A_r; \sigma; \underline{w}_{r-1}) = \left(\bigotimes_{k=1}^r A_k \right) G_{q\underline{1}_r}(\sigma; \underline{w}_{r-1}) \left(\bigotimes_{k=1}^r A_k \right)^t.$$

It follows that

$$(7.3) \quad \begin{aligned} H_{q\underline{1}_r}(I, \dots, I; \sigma; \underline{w}_{r-1}) &= G_{q\underline{1}_r}(\sigma; \underline{w}_{r-1}) \\ H_{q\underline{1}_r}(a_1 A_1, \dots, a_r A_r; \sigma; \underline{w}_{r-1}) &= \left(\prod_{k=1}^r a_k^2 \right) H_{q\underline{1}_r}(A_1, \dots, A_r; \sigma; \underline{w}_{r-1}) \end{aligned}$$

and

$$(7.4) \quad H_{\underline{1}_r}(a_1, \dots, a_r; \sigma; \underline{w}_{r-1}) = \prod_{k=1}^r a_k^2$$

for scalars $a_1, \dots, a_r$. If $A_i = 0$ for some i , $1 \leq i \leq r$, then

$$(7.5) \quad H_{q\underline{1}_r}(A_1, \dots, A_r; \sigma; \underline{w}_{r-1}) = 0$$

We shall need two special cases of these matrices:

$$H_{q\underline{1}_r}(r; A; \sigma; \underline{w}_{r-1}) \equiv H_{q\underline{1}_r}(A, \dots, A; \sigma; \underline{w}_{r-1})$$

and

$$H_{q\underline{1}_r}(1, r-1; C, A; i; \sigma; \underline{w}_{r-1}) \equiv H_{q\underline{1}_r}(\underbrace{A, \dots, A}_{i-1}, C, \underbrace{A, \dots, A}_{r-i}; \sigma; \underline{w}_{r-1}),$$

$i=1, \dots, r$. From (7.2) we infer the following alternative expressions for these matrices

$$H_{q\underline{1}_r}(r; A; \sigma; \underline{w}_{r-1}) = A^{\langle r \rangle} G_{q\underline{1}_r}(\sigma; \underline{w}_{r-1}) A^{\langle r \rangle t}$$

and

$$\begin{aligned}
 H_{q \underline{1}_r}(1, r-1: C, A: i+1; \sigma; \underline{w}_{r-1}) &= (A^{\langle \min(\sigma(i+1)_{\underline{1}_i}, \sigma(i+1)_{\underline{1}_{i+1}})) - 1 \rangle} \\
 &\otimes C^{\langle 1 - \underline{1}_i t_{i+1} \rangle} \otimes A^{\langle |\sigma(i) - \sigma(i+1)| + \underline{1}_i t_{i+1} - 1 \rangle} \otimes C^{\langle t_{\underline{1}_i} \underline{1}_{i+1} \rangle} \\
 &\otimes A^{\langle r+1 - t_{\underline{1}_i} \underline{1}_{i+1} - \max(\sigma(i+1)_{\underline{1}_i}, \sigma(i+1)_{\underline{1}_{i+1}}) \rangle}) G_{q \underline{1}_r}(\sigma; \underline{w}_{r-1}) \\
 &\times (A^{\langle \min(\sigma(i+t)_{\underline{1}_i}, \sigma(i+t)_{\underline{1}_{i+1}})) - 1 \rangle} \otimes C^{\langle 1 - t_{\underline{1}_i} \underline{1}_{i+1} \rangle} \\
 &\otimes A^{\langle |\sigma(i) - \sigma(i+1)| + t_{\underline{1}_i} \underline{1}_{i+1} - 1 \rangle} \otimes C^{\langle \underline{1}_i t_{i+1} \rangle} \\
 &\otimes A^{\langle r+1 - \underline{1}_i t_{i+1} - \max(\sigma(i+t)_{\underline{1}_i}, \sigma(i+t)_{\underline{1}_{i+1}}) \rangle})^t
 \end{aligned}$$

for $i=0, 1, \dots, r-1$, where $\underline{1}_i = 1$ if $w_i = 1$ and zero otherwise and where $w_0 = w_r = t$, $\sigma(0) = 1$ and $\sigma(r) = r$.

We define "averages"

$$H_q(1:A) = AA^t$$

and

$$H_{q \underline{1}_r}(r:A) = \sum_{\sigma \in S_{\underline{1}_{r-1}}} \sum_{\underline{w}_{r-1} \in \{1, t\}^{r-1}} H_{q \underline{1}_r}(r:A; \sigma; \underline{w}_{r-1}) / ((r-1)! 2^{r-1})$$

for $r=2, 3, \dots$,

$$H_q(1, 0: C, A: 1) = CC^t$$

and

$$\begin{aligned}
 H_{q \underline{1}_r}(1, r-1: C, A: i) \\
 = \sum_{\sigma \in S_{\underline{1}_{r-1}}} \sum_{\underline{w}_{r-1} \in \{1, t\}^{r-1}} H_{q \underline{1}_r}(1, r-1: C, A: i; \sigma; \underline{w}_{r-1}) / ((r-1)! 2^{r-1})
 \end{aligned}$$

for $r=2, 3, \dots$, and $i=1, \dots, r$. We also define

$$\Gamma_{q \underline{1}_r}(r) = \sum_{\sigma \in S_{\underline{1}_{r-1}}} \sum_{\underline{w}_{r-1} \in \{1, t\}^{r-1}} G_{q \underline{1}_r}(\sigma; \underline{w}_{r-1}) / 2^{r-1}.$$

which will play the same role for the moments and cumulants of the Wishart distribution as does the gammafunction for the moments and cumulants of the chi-square distribution.

7.3. Some matrix derivatives

By aid of Lemma 5.1 we obtain the following results.

THEOREM 7.1. For $-\infty < \alpha < \infty$ and for matrices A, B, C, D, E, F, G and H which do not depend on the elements of the matrix T , and such that the operations are well defined, the following results hold.

- (i) $D_T \otimes |I+ETF+GT^tH|^\alpha$
 $= \alpha |I+ETF+GT^tH|^\alpha (E^t(I+ETF+GT^tH)^{-t}F^t + H(I+ETF+GT^tH)^{-1}G).$
- (ii) $D_T \otimes \ln(|I+ETF+GT^tH|^\alpha) = \alpha (E^t(I+ETF+GT^tH)^{-t}F^t + H(I+ETF+GT^tH)^{-1}G).$
- (iii) $D_T \otimes \text{tr}((ATB+CT^tD)(I+ETF+GT^tH)^{-1})$
 $= A^t(I+ETF+GT^tH)^{-t}B^t + D(I+ETF+GT^tH)^{-1}C$
 $- E^t(I+ETF+GT^tH)^{-t}(ATB+CT^tD)^t(I+ETF+GT^tH)^{-t}F^t$
 $- H(I+ETF+GT^tH)^{-1}(ATB+CT^tD)(I+ETF+GT^tH)^{-1}G.$
- (iv) $D_T \otimes \exp(\text{tr}((ATB+CT^tD)(I+ETF+GT^tH)^{-1}))$
 $= \exp(\text{tr}((ATB+CT^tD)(I+ETF+GT^tH)^{-1}))$
 $\times (A^t(I+ETF+GT^tH)^{-t}B^t + D(I+ETF+GT^tH)^{-1}C$
 $- E^t(I+ETF+GT^tH)^{-t}(ATB+CT^tD)^t(I+ETF+GT^tH)^{-t}F^t$
 $- H(I+ETF+GT^tH)^{-1}(ATB+CT^tD)(I+ETF+GT^tH)^{-1}G).$

PROOF. (i): By using the chain-rule we have

$$\begin{aligned} d_T \otimes |I+ETF+GT^tH|^\alpha &= (d_{|I+ETF+GT^tH|} \otimes |I+ETF+GT^tH|^\alpha) \\ &\quad \times (d_{I+ETF+GT^tH} \otimes |I+ETF+GT^tH|) (d_T \otimes (I+ETF+GT^tH)) \\ &= \alpha |I+ETF+GT^tH|^{\alpha-1} |I+ETF+GT^tH| (\text{vec}((I+ETF+GT^tH)^{-1}))^t \\ &\quad \times ((E \otimes F^t) + (G \otimes H^t)) \\ &= \alpha |I+ETF+GT^tH|^\alpha (\text{vec}(E^t(I+ETF+GT^tH)^{-t}F^t + H(I+ETF+GT^tH)^{-1}G))^t \end{aligned}$$

The result now follows from (5.11).

(ii): From the chain-rule we infer that

$$d_T \otimes \ln(|I+ETF+GT^tH|^\alpha) = |I+ETF+GT^tH|^{-\alpha} (d_T \otimes |I+ETF+GT^tH|^\alpha)$$

and hence the result follows from (i).

(iii): By using the chain-rule, product-rule and inverse-rule we obtain

$$\begin{aligned}
 d_T &\square \operatorname{tr}((ATB+CT^tD)(I+ETF+GT^tH)^{-1}) \\
 &= (\operatorname{vec}(I))^t((I \otimes (I+ETF+GT^tH)^{-t}))(d_T \square (ATB+CT^tD)) \\
 &\quad -((ATB+CT^tD) \otimes I)((I+ETF+GT^tH)^{-1} \otimes (I+ETF+GT^tH)^{-t}) \\
 &\quad \times (d_T \square (I+ETF+GT^tH)) \\
 &= (\operatorname{vec}(I))^t((I \otimes (I+ETF+GT^tH)^{-t})(A \otimes B^t + C \otimes D^t) \\
 &\quad -((ATB+CT^tD)(I+ETF+GT^tH)^{-1} \otimes (I+ETF+GT^tH)^{-t}) \\
 &\quad \times (E \otimes F^t + G \otimes H^t)) \\
 &= (\operatorname{vec}(A^t(I+ETF+GT^tH)^{-t}B^t + D(I+ETF+GT^tH)^{-1}C \\
 &\quad -E^t(I+ETF+GT^tH)^{-t}(ATB+CT^tD)^t(I+ETF+GT^tH)^{-t}F^t \\
 &\quad -H(I+ETF+GT^tH)^{-1}(ATB+CT^tD)(I+ETF+GT^tH)^{-1}G))^t.
 \end{aligned}$$

The result now follows from (5.11).

(iv): From the chain-rule we have

$$\begin{aligned}
 d_T &\square \exp(\operatorname{tr}((ATB+CT^tD)(I+ETF+GT^tH)^{-1})) \\
 &= \exp(\operatorname{tr}((ATB+CT^tD)(I+ETF+GT^tH)^{-1})) \\
 &\quad \times (d_T \square \operatorname{tr}((ATB+CT^tD)(I+ETF+GT^tH)^{-1}))
 \end{aligned}$$

and hence the result follows from (iii). $\square$

As a result of this theorem we obtain the following corollary.

COROLLARY 7.1. For matrices Ω and V which do not depend on T and such that the operations are well defined, we have

$$\begin{aligned}
 (i) \quad D_T &\otimes (\ln(|I-V(T+T^t)|^{-\alpha}) + \operatorname{tr}(\Omega(T+T^t)(I-V(T+T^t))^{-1})) \\
 &= (\alpha V^t + (I+V^t(I-V(T+T^t))^{-t}(T+T^t))\Omega^t)(I-V(T+T^t))^{-t} \\
 &\quad + (I-V(T+T^t))^{-1}(\alpha V + \Omega(I+(T+T^t)(I-V(T+T^t))^{-1}V).
 \end{aligned}$$

(ii) If in addition $\sum_{j=0}^n (V(T+T^t))^j$ and $\sum_{j=0}^n ((T+T^t)V)^j$ converges as $n \rightarrow \infty$ then

$$\begin{aligned}
 D_T &\otimes (\ln(|I-V(T+T^t)|^{-\alpha}) + \operatorname{tr}(\Omega(T+T^t)(I-V(T+T^t))^{-1})) \\
 &= (\alpha V^t + (I-(T+T^t)V)^{-t}\Omega^t)(I-V(T+T^t))^{-t} \\
 &\quad + (I-\bar{V}(T+T^t))^{-1}(\alpha V + \Omega(I-(T+T^t)V)^{-1}).
 \end{aligned}$$

PROOF. (i): By combining Theorem 7.1 (ii) and (iii) for $B = D = F = H = I$, $E = G = -V$, $A = C = \Omega$ and with α replaced by $-\alpha$, the result immediately follows.

(ii) If $\sum_{j=0}^n (V(T+T^t))^j$ converges as $n \rightarrow \infty$, then

$$(7.6) \quad (I - V(T+T^t))^{-1} = \sum_{j=0}^{\infty} (V(T+T^t))^j$$

and hence

$$I + (T+T^t)(I - V(T+T^t))^{-1}V = \sum_{j=0}^{\infty} ((T+T^t)V)^j = (I - (T+T^t)V)^{-1}.$$

where the last equation follows from that $\sum_{j=0}^n ((T+T^t)V)^j$ converges to $(I - (T+T^t)V)^{-1}$. The result now follows from (i). $\square$

The following theorem is the one on which the derivation of moments and cumulants mainly rests.

THEOREM 7.2. Let A_j , $j=1, \dots, r$ and X be q -square matrices and let $\underline{w}_{r-1} = (w_1, \dots, w_{r-1})$ be a $(r-1)$ -tuple of l 's and t 's. Then

$$\begin{aligned} D_X^{<r-1>} \otimes (A_1 A_1^t X^{w_1} A_2 A_2^t \cdots A_{r-1} A_{r-1}^t X^{w_{r-1}} A_r A_r^t) \\ = \sum_{\sigma \in S_{\underline{1}_{r-1}}} H_{q \underline{1}_r} (A_1, \dots, A_r; \sigma; \underline{w}_{r-1}) \end{aligned}$$

PROOF. Let $X = \sum_{i=1}^q \sum_{j=1}^q x_{i,j} E_{i,j}$. Then

$$\begin{aligned} A_1 A_1^t X^{w_1} A_2 A_2^t \cdots A_{r-1} A_{r-1}^t X^{w_{r-1}} A_r A_r^t \\ = \sum^* \left(\sum_{k=1}^{r-1} b_{i_k, j_k}^{(k)} x_{w_k(j_k, i_{k+1})} \right) b_{i_r, j_r}^{(r)} E_{i_1, j_r} \end{aligned}$$

where $A_k A_k^t = \sum_{i=1}^q \sum_{j=1}^q b_{i,j}^{(k)} E_{i,j}$, the symbol $w_k(j_k, i_{k+1})$ stands for (j_k, i_{k+1}) if $w_k=1$ but (i_{k+1}, j_k) if $w_k=t$, and where $\sum^*$ extends over all integer values of i_k and j_k , $k=1, \dots, r$ in $\{1, 2, \dots, r\}$.

Since for arbitrary $C_k = \sum_{i=1}^{m_k} \sum_{j=1}^{n_k} c_{i,j}^{(k)} E_{i,j}$, $k=1, \dots, r$

$$C_1 \otimes \dots \otimes C_r = \sum_{\underline{i}_r \in I(\underline{m}_r)} \sum_{\underline{j}_r \in I(\underline{n}_r)} \left(\prod_{k=1}^r c_{i_k, j_k}^{(k)} \right) E_{p(\underline{i}_r; \underline{m}_r), p(\underline{j}_r; \underline{n}_r)}$$

cf. Holmquist (1985, Theorem 3.3 (i)), we see that

$$(7.7) \quad D_X^{<r-1>} \otimes (A_1 A_1^t X^{w_1} A_2 A_2^t \dots A_{r-1} A_{r-1}^t X^{w_{r-1}} A_r A_r^t) \\ = \Sigma^{**} \Sigma^* \partial^{r-1} \left(\left(\prod_{k=1}^{r-1} b_{i_k, j_k}^{(k)} x_{w_k(j_k, i_{k+1})} \right) b_{i_r, j_r}^{(r)} / \left(\prod_{k=1}^{r-1} \partial x_{\alpha_k, \beta_k} \right) \right) \\ \times E_{p(\alpha_1, \dots, \alpha_{r-1}, i_1), p(\beta_1, \dots, \beta_{r-1}, j_r)}$$

where Σ^{**} extends over all integer values of α_k and β_k , $k=1, \dots, r-1$, in $\{1, \dots, r\}$. Further, $\partial^{r-1} \left(\prod_{k=1}^{r-1} x_{w_k(j_k, i_{k+1})} \right) / \left(\prod_{k=1}^{r-1} \partial x_{\alpha_k, \beta_k} \right)$ is equal to one as soon as

$$(\alpha_k, \beta_k) = w_{\sigma^{-1}(k)}^{j_{\sigma^{-1}(k)}, i_{\sigma^{-1}(k)+1}}$$

for some $\sigma \in S_{\underline{1}_{r-1}}$, but equal to zero otherwise. Hence the right hand side of (7.7) can be written

$$(7.8) \quad \sum_{\sigma \in S_{\underline{1}_{r-1}}} \Sigma^* \left(\prod_{k=1}^r b_{i_k, j_k}^{(k)} \right) E_{p(j_1'', \dots, j_r''), p(i_1^w, \dots, i_r^w)}$$

where (j_k'', i_k^w) is equal to (j_k'', i_k) if $w_{\sigma^{-1}(k)} = 1$ but equal to (i_k'', j_k) if $w_{\sigma^{-1}(k)} = t$, for $k=1, \dots, r-1$ and where $(j_r'', i_r^w) = (i_1, j_r)$.

Next we show that the summation over $S_{\underline{1}_{r-1}}$ of $H_{q\underline{1}_r}(A_1, \dots, A_r; \sigma; \underline{w}_{r-1})$ can be written in the form (7.8). By using (7.2) we can write this summation

$$(7.9) \quad \sum_{\sigma \in S_{\underline{1}_{r-1}}} \left(\sum_1 \left(\prod_{k=1}^r a_{j_k'', \alpha_k}^{(k'')} \right) E_{p(j_1^w, \dots, j_r^w), p(\alpha_1, \dots, \alpha_r)} \right) \\ \times \left(\sum_2 E_{p(\delta_1'', \dots, \delta_r''), p(\delta_1, \dots, \delta_r)} \right) \\ \times \left(\sum_3 \left(\prod_{k=1}^r a_{i_k^w, \beta_k}^{(k')} \right) E_{p(\beta_1, \dots, \beta_r), p(i_1^w, \dots, i_r^w)} \right)$$

where the summations Σ_1 , Σ_2 and Σ_3 extend over all integer values from the set $\{1, 2, \dots, q\}$ of the index variables in the corresponding summation, and $a_{i,j}^{(k)}$ is the (i,j) 'th coordinate of A_k . By performing

the multiplication in (7.9), this expression reduces to

$$(7.10) \quad \sum_{\sigma \in S_{\underline{r-1}}} \sum_{k=1}^r a_{j_{k''}, \delta_{k''}}^{(k'')} a_{i_{k'}, \delta_{k'}}^{(k')} E_{p(j_{1''}, \dots, j_{r''}), p(i_{1'}, \dots, i_{r'})}.$$

Here the second summation extends over all integer values from the set $\{1, 2, \dots, q\}$ of the variables i_k, j_k and $\delta_k, k=1, \dots, r$.

We note that

$$a_{j_{k''}, \delta_{k''}}^{(k'')} a_{i_{k'}, \delta_{k'}}^{(k')} = a_{j_{\sigma^{-1}(k)}, \delta_{\sigma^{-1}(k)}}^{(\sigma^{-1}(k))} a_{i_{\sigma^{-1}(k)+1}, \delta_{\sigma^{-1}(k)+1}}^{(\sigma^{-1}(k)+1)}$$

for $k=1, \dots, r-1$ and hence we have

$$(7.11) \quad \prod_{k=1}^r a_{j_{k''}, \delta_{k''}}^{(k'')} a_{i_{k'}, \delta_{k'}}^{(k')} = \prod_{k=1}^r a_{j_k, \delta_k}^{(k)} a_{i_k, \delta_k}^{(k)}$$

independently of $\underline{r-1}$. Further, if we sum terms given by the right hand side of (7.11) for all the possible values of $\delta_1, \dots, \delta_r$, the result is

$\prod_{k=1}^r b_{i_k, j_k}^{(k)}$. This shows that the expression (7.10) can be reduced to (7.8), which thus proves the theorem. $\square$

The result given in Theorem 7.2 is a generalization of

$$(7.12) \quad D_X \otimes X = \text{vec}(I_q) \text{vec}(I_q)^t$$

and

$$(7.13) \quad D_X \otimes X^t = T_{q,q}$$

for q -square matrices X ; cf. e.g. MacRae (1974, Theorem 1). The results

(7.12) and (7.13) are obtained from Theorem 7.2 by taking $r=2$ and A_1

$= A_2 = I_q$ and noting that $G_{q,q}(1;1) = \text{vec}(I_q) \text{vec}(I_q)^t$ and $G_{q,q}(1;t) = T_{q,q}$

where 1 is the identity permutation in S_1 . It is straightforward to

generalize Theorem 7.2 to non-square matrices X so that it generalizes

Theorem 1 of MacRae (1974). For our application we shall however not need

this generalization.

7.4. Cumulants and moments of the Wishart distribution

We are now ready to calculate the cumulants of arbitrary order.

THEOREM 7.3. The r 'th cumulant of the r.v. having m.g.f. given by (7.1) is

$$\kappa_{r,X} = \alpha(r-1)! H_{q_{1-r}}(r; B^{\frac{1}{2}}) + (r-1)! \sum_{i=1}^r H_{q_{1-r}}(1, r-1; \Lambda^{\frac{1}{2}}, B^{\frac{1}{2}}; i)$$

where $B^{\frac{1}{2}}(B^{\frac{1}{2}})^t = B$ and $\Lambda^{\frac{1}{2}}(\Lambda^{\frac{1}{2}})^t = \Lambda$.

PROOF. According to (6.2) the r 'th cumulant is equal to the r 'th derivative with respect to T at zero of

$$\ln(|I - V(T+T^t)|^{-\alpha}) + \text{tr}(\Omega(T+T^t)(I - V(T+T^t))^{-1})$$

where $V = B/2$ and $\Omega = \Lambda/2$. From Corollary 7.1 (ii) we infer that

$$\begin{aligned} (7.14) \quad & D_T^{<r>} \otimes (\ln(|I - V(T+T^t)|^{-\alpha}) + \text{tr}(\Omega(T+T^t))^{-1}) \\ &= D_T^{<r-1>} \otimes (\alpha V^t (I - V(T+T^t))^{-t} + \alpha (I - V(T+T^t))^{-1} V \\ &\quad + (I - V(T+T^t))^{-t} \Omega^t (I - V(T+T^t))^{-t} \\ &\quad + (I - V(T+T^t))^{-1} \Omega (I - (T+T^t)V)^{-1}) \end{aligned}$$

for T sufficiently close to the zero matrix so that $\sum_{j=0}^n (V(T+T^t))^j$ converges as n tends to infinity. By using the expansion (7.6) we write (7.14) as

$$(7.15) \quad D_T^{<r-1>} \otimes (2\alpha \sum_{j=0}^{\infty} V((T+T^t)V)^j + 2 \sum_{j=0}^{\infty} \sum_{i=0}^j (V(T+T^t))^i \Omega ((T+T^t)V)^{j-i})$$

where we have also used the fact that both V and Ω are symmetric.

In fact, when T is equated to zero, only terms in (7.15) that are $(r-1)$ -monomials in $(T+T^t)$ contribute, since lower order terms vanish and higher order terms are equated to zero when T is equated to zero.

Hence we conclude that

$$\kappa_{r,X} = D_T^{<r-1>} \otimes (2\alpha V((T+T^t)V)^{r-1} + 2 \sum_{i=0}^{r-1} (V(T+T^t))^i \Omega ((T+T^t)V)^{r-1-i} \Big|_{T=0}.$$

By using Theorem 7.2 we then obtain

$$\begin{aligned}
 \kappa_{r,X} &= D_T^{<r-1>} \otimes \left(2 \sum_{\underline{w}_{r-1} \in \{1,t\}^{r-1}} (\alpha V_T^{w_1} V_T^{w_2} \dots V_T^{w_{r-1}} V \right. \\
 &\quad \left. + \sum_{i=0}^{r-1} V_T^{w_1} V \dots V_T^{w_i} \Omega_T^{w_{i+1}} V \dots V_T^{w_{r-1}} V) \right) \Big|_{T=0} \\
 &= 2 \sum_{\underline{w}_{r-1} \in \{1,t\}^{r-1}} \left(\sum_{\sigma \in S_{\frac{1}{2}r-1}} \alpha H_{q \frac{1}{2}r} (r: V^{\frac{1}{2}}; \sigma; \underline{w}_{r-1}) \right. \\
 &\quad \left. + \sum_{i=0}^{r-1} H_{q \frac{1}{2}r} (1, r-1: \Omega^{\frac{1}{2}}; V^{\frac{1}{2}}; i+1; \sigma; \underline{w}_{r-1}) \right) \\
 &= 2(2^{r-1}(r-1)! \alpha H_{q \frac{1}{2}r} (r: V^{\frac{1}{2}}) \\
 &\quad + \sum_{i=1}^r 2^{r-1}(r-1)! H_{q \frac{1}{2}r} (1, r-1: \Omega^{\frac{1}{2}}, V^{\frac{1}{2}}; i)) \\
 &= \alpha(r-1)! H_{q \frac{1}{2}r} (r: B^{\frac{1}{2}}) + (r-1)! \sum_{i=1}^r H_{q \frac{1}{2}r} (1, r-1: \Lambda^{\frac{1}{2}}, B^{\frac{1}{2}}; i);
 \end{aligned}$$

in the last step we have used (7.3) for $B^{\frac{1}{2}} = V^{\frac{1}{2}} 2^{\frac{1}{2}}$ and $\Lambda^{\frac{1}{2}} = \Omega^{\frac{1}{2}} 2^{\frac{1}{2}}$. $\square$

EXAMPLE 7.1. The first two cumulants are

$$\begin{aligned}
 \kappa_{1,X} &= \alpha B + \Lambda \\
 \kappa_{2,X} &= (\alpha/2) B^{\frac{1}{2} < 2 >} (T_{q,q} + \text{vec} I_q (\text{vec} I_q)^t) B^{\frac{1}{2} < 2 > t} + \\
 &\quad + (1/2) ((\Lambda^{\frac{1}{2}} \otimes B^{\frac{1}{2}}) T_{q,q} (B^{\frac{1}{2}} \otimes \Lambda^{\frac{1}{2}})^t + B^{\frac{1}{2} < 2 >} \text{vec} I_q (\text{vec} I_q)^t \Lambda^{\frac{1}{2} < 2 > t} \\
 &\quad + (B^{\frac{1}{2}} \otimes \Lambda^{\frac{1}{2}}) T_{q,q} (\Lambda^{\frac{1}{2}} \otimes B^{\frac{1}{2}})^t + \Lambda^{\frac{1}{2} < 2 >} \text{vec} I_q (\text{vec} I_q)^t B^{\frac{1}{2} < 2 > t}) \\
 &= (\alpha/2) (T_{q,q} B^{\frac{1}{2} < 2 >} + \text{vec} B (\text{vec} B)^t) \\
 &\quad + (1/2) (T_{q,q} (B \otimes \Lambda + \Lambda \otimes B) + \text{vec} B (\text{vec} \Lambda)^t + \text{vec} \Lambda (\text{vec} B)^t).
 \end{aligned}$$

By setting $B = 2V$ and $\alpha = n/2$, we obtain the first two cumulants of the $W_q(n, V, \Lambda)$ distribution. These coincide with the first two central moments $E(X)$ and $E((X-E(X))^{\langle 2 \rangle}) = E(X^{\langle 2 \rangle}) - E(X)^{\langle 2 \rangle}$ (cf. Theorem 4.1 (iii)). By comparison to the result of Magnus and Neudecker (1979), we see that the expressions for the mean coincide but not the expressions for the dispersion. This is because they define the dispersion matrix of a random matrix to be the dispersion matrix of the vectorized random matrix. $\square$

As a special case we have the central distribution, where $\Lambda = 0$.

COROLLARY 7.2. The cumulants of a symmetric random matrix with m.g.f. given by $|I-BT|^{-\alpha}$ where $\alpha > (q-1)/2$ or $\alpha = 1/2, 1, 3/2, \dots, (q-1)/2$, B is q -square p.s.d. and T is symmetric, are

$$\kappa_r = \alpha(r-1)! H_{q1_r}(r; B^{\frac{1}{2}}) = \alpha B^{\frac{1}{2}\langle r \rangle} \Gamma_{q1_r}(r) B^{\frac{1}{2}\langle r \rangle t}, \quad r=1, 2, \dots \quad \square$$

This result follows from Theorem 7.3 and (7.5).

Another special case is when $q=1$.

COROLLARY 7.3. The cumulants of a univariate random variable with m.g.f. given by $(1-\beta t)^{-\alpha} \exp((\lambda t(1-\beta t)^{-1})$ where $\alpha > 0$, $\beta \geq 0$ and $\lambda \geq 0$, are

$$\kappa_r = \alpha(r-1)! \beta^r + r! \lambda \beta^{r-1}, \quad r=1, 2, \dots \quad \square$$

This result follows from Theorem 7.3 and (7.4). For $\beta = 2$ and $\alpha = n/2$ these are the cumulants of the non-central χ^2 -distribution, $\chi^2(n, \lambda)$, with n degrees of freedom and non-centrality parameter λ (cf. e.g. Johnson and Kotz (1970, p. 134).

The calculations of the r 'th moment about zero by means of differentiating the moment generating function, can be obtained in a similar manner. Once the cumulants are obtained it is, however, simpler to use Theorem 4.1 to obtain the following result.

THEOREM 7.4. With the same notation as in Theorem 4.1 it holds that the r 'th moment about zero of a symmetric random matrix having m.g.f. given by (7.1) is

$$\begin{aligned} \mu_{r,X} = & \sum_{j_r}^r (1/\prod_{i=1}^r j_i! i^{j_i}) \left(\sum_{\sigma \in S_{\frac{1}{2}r}} P_{q1_r}(\sigma) \left(\bigotimes_{i=1}^r (\alpha H_{q1_i}(i; B^{\frac{1}{2}}) \right. \right. \\ & \left. \left. + \sum_{k=1}^i H_{q1_i}(1, i-1; \Lambda^{\frac{1}{2}}, B^{\frac{1}{2}}; k) \right)^{\langle j_i \rangle} P_{q1_r}(\sigma)^t \right). \end{aligned}$$

PROOF. The result is obtained by substituting the result of Theorem 7.3

into Theorem 4.1 (i) and using (2.5) and (6.3). $\square$

As a special case we have the moments of the central Wishart distribution.

COROLLARY 7.4. The r 'th moment of a symmetric random matrix with m.g.f.

given by $|I-BT|^{-\alpha}$ where $\alpha > (q-1)/2$ or $\alpha = 1/2, 1, 3/2, \dots, (q-1)/2$,

B is q -square p.s.d. and T is symmetric, is

$$\mu_{r,X} = \sum_{j_1, \dots, j_r} \frac{\alpha^{j_1 + \dots + j_r}}{\prod_{i=1}^r j_i! (i!)^{j_i}} \times B^{\frac{1}{2}\langle r \rangle} \left(\sum_{\sigma \in S_{\underline{1-r}}} P_{q\underline{1-r}}(\sigma) \left(\bigotimes_{i=1}^r \Gamma_{q\underline{1-i}}(i)^{\langle j_i \rangle} \right) P_{q\underline{1-r}}(\sigma)^t \right) B^{\frac{1}{2}\langle r \rangle}. \quad \square$$

Remark 7.1. It should be noted that there is an alternative way to obtain

the moments of the Wishart distribution which is similar to the approach

used by Magnus and Neudecker (1979) to obtain the first two moments. Let

$X = \sum_{i=1}^n x_i x_i^t$ where $x_i \in N_q(m_i, V)$ and x_i, x_j are independent for $i \neq j$.

Since $X \in W_q(n, V, \sum_{i=1}^n m_i m_i^t)$ the moments of the Wishart distribution can

easily be calculated from the multinomial expansion (2.6) and the moments

of the multivariate normal distribution:

$$\begin{aligned} \mu_{r,X} &= E(X^{\langle r \rangle}) \\ &= \sum_{\underline{r-n}} \sum_{\sigma \in S_{\underline{r-n}}} P_{q\underline{1-r}}(\sigma) (E(Y_1^{\langle r_1 \rangle} \otimes \dots \otimes Y_n^{\langle r_n \rangle})) P_{q\underline{1-r}}(\sigma)^t \\ &= \sum_{\underline{r-n}} \sum_{\sigma \in S_{\underline{r-n}}} P_{q\underline{1-r}}(\sigma) (\mu_{r_1, Y_1} \otimes \dots \otimes \mu_{r_n, Y_n}) P_{q\underline{1-r}}(\sigma)^t. \end{aligned}$$

Here $Y_i = x_i x_i^t$; the last equality follows from the independence of the

Y_i 's. Since the elements of μ_{k, Y_i} is equal to a proper rearrangement of

the elements of μ_{2k, x_i} , more precisely $\text{vec}_{kq, kq}(\mu_{k, Y_i}) = T_{kq, kq} \mu_{2k, x_i}$,

the moments of X can thus be obtained from the even moments of the x_i 's.

These latter moments can be obtained by aid of Theorem 4.1 and the fact

that $\kappa_{1, x_i} = m_i$, $\kappa_{2, x_i} = \text{vec}(V)$ and $\kappa_{r, x_i} = 0$ for $r > 2$. $\square$

REFERENCES

- Aitken, A.C. (1931). Some applications of generating functions to normal frequency. *Quart. Journ. Math. Oxford ser. 2*, 130-135.
- Anderson, T.W. (1958). *An Introduction to Multivariate Statistical Analysis*. John Wiley and Sons, New York.
- Brillinger, O.R. (1975). *Time Series, Data Analysis and Theory*. Holt, Rinehart and Winston, New York.
- DeWaal, D.J. and Nel, D.G. (1973). On some expectations with respect to Wishart matrices. *South African Statist. J.* 7, 61-67.
- Graybill, F.A. (1969). *Introduction to Matrices with Applications in Statistics*. Wadsworth Publishing Co., Belmont, California.
- Hald, A. (1981). T.N. Thiele's contribution to statistics. *Internat. Statist. Rev.* 49, 1-20.
- Henderson, H.V. and Searle, S.R. (1981). The Vec-permutation matrix, the Vec operator and Kronecker products: A review. *Linear and Multilinear Algebra* 9, 271-288.
- Hille, E. (1948). *Functional Analysis and Semi-groups*. Amer. Math. Soc. Coll. Publ., New York.
- Holmquist, B. (1985). The direct product permuting matrices. *Linear and Multilinear Algebra* 17, 117-141.
- Johnson, N.L. and Kotz, S. (1970). *Distributions in Statistics: Continuous Univariate Distributions - 2*. Houghton Mifflin Co., Boston.
- Kendall, M.G. and Stuart, A. (1976). *The Advanced Theory of Statistics*, Vol 1. Fourth edition. Charles Griffin and Co., London.
- Leonov, V.P. and Shiryaev, A.N. (1959). On a method of calculation of semi-invariants. *Theor. Prob. Appl.* 4, 319 -329.
- MacRae, E.C. (1974). Matrix derivatives with an application to an adaptive linear decision problem. *Ann. Statist.* 2, 337-346.
- Magnus, J.R. and Neudecker, H. (1979). The commutation matrix: Some properties and applications. *Ann. Statist.* 7, 381-394.

- Marcus, M. (1973). Finite Dimensional Multilinear Algebra, Part I. Marcel Dekker, New York.
- McCullagh, P. (1984). Tensor notation and cumulants of polynomials. *Biometrika* 71, 461-476.
- McCulloch, C.E. (1982). Symmetric matrix derivatives with applications. *J. Amer. Statist. Assoc.* 77, 679-682.
- McDonald, R.P. and Swaminathan, H. (1973). A simple matrix calculus with applications to multivariate analysis. *General Systems* 18, 37-54.
- Mourier, E. (1953). Eléments aléatoires dans un espace de Banach. *Ann. Inst. H. Poincaré* 13, 161-244.
- Nel, D.G. (1980). On matrix differentiation in statistics. *South African Statist. J.* 14, 137-193.
- Neudecker, H. (1968). The Kronecker matrix product and some of its applications in econometrics. *Statist. Neerlandica* 22, 69-82.
- Neudecker, H. and Wansbeck, T. (1983). Some aspects on commutation matrices, with statistical applications. *Canad. J. Statist.* 11, 221-231.
- Patil, G.P. and Boswell, M.T. (1970). A characteristic property of the multivariate normal distribution and some of its applications. *Ann. Math. Statist.* 41, 1970-1977.
- Pettis, B.J. (1938). Integration in vector spaces. *Trans. Amer. Math. Soc.* 44, 277-309.
- Plackett, R.L. (1954). A reduction formula for normal multivariate integrals. *Biometrika* 41, 351-360.
- Rao, C.R. (1973). *Linear Statistical Inference and its Applications*. Second edition. John Wiley and Sons, New York.
- Rao, C.R. and Mitra, S.K. (1971). *Generalized Inverse of Matrices and its Applications*. John Wiley and Sons, New York.
- Rosen, D. von (1984). Maximum likelihood estimates in multivariate linear normal models with special references to the growth curve model.

Inst. of Actuar. Mathematics and Math. Statist., Univ. of Stockholm,
Techn. Report No. 135, 1-149.

Rogers, G.S. (1980). Matrix Derivatives. Marcel Dekker, New York.

Speed, T.P. (1983). Cumulants and partition lattices. Austral. J. Statist.
25, 378-388.

Swaminathan, H. (1976). Matrix calculus for functions of partitioned
matrices. General Systems 21, 95-99.

Thiele, T.N. (1889). Forelaesninger over Almindlig Iagttagelselaere: Sand-
syldighetsregning og mindste Kvadraters Methode. Reitzel, København.
(Original not seen, cited in Hald (1981).)

Thiele, T.N. (1899). Om iagttagelselaerens halvinvarianter. Oversigt
Vidensk. Selsk. Forhandlinger, 135-141.

Tracy, D.S. and Dwyer, P.S. (1969). Multivariate maxima and minima with
matrix derivatives. J. Amer. Statist. Assoc. 64, 1576-1594.

Vakhania, N.N. (1981). Probability Distributions on Linear Spaces.
North Holland, New York.

